The Virtual Stages of Hate: Using Goffman's Work to Conceptualise the Motivations for Online Hate

Introduction

It was originally predicted that reduced social cues on the Internet would result in less meaningful interaction meaning that social relationships online could not be significantly developed (Hine, 2012). With the benefit of hindsight, this predication could not be further from the truth as Web 2.0 and social media platforms, in particular, have revolutionised the way contemporary audiences communicate (Nakamura, 2008). Social media dominates our daily experiences as we use these platforms to construct our virtual identities, create and develop friendships, locate job opportunities, play games, and consume the latest news (Farrington et al. 2015; Lind, 2019; Moore et al. 2017). Merunkova and Slerka (2019: 244-5) state that 'Three-quarters of Facebook users log in daily' while 91 percent of teenagers go 'online every day, which is an indication of the importance of cyberspace in their lives'.

Hine (2012: 3) suggests that 'the internet is both a hugely significant social phenomenon of our time in itself and, in turn, a fascinating field site for social science research of all kinds'. Therefore, research into online interaction and behaviour is of paramount importance as offline inequalities are reflected in cyberspace (boyd, 2011; Lind, 2019). Hence, investigating online communities and experiences helps us, for example, critically understand the social, cultural and economic impact of the digital divide (West and Thakore, 2013) and 'white flight' (boyd, 2011). It also allows us to comprehend the correlation between online communication and offline behaviours (Brown, 2009; Keum and Miller, 2018; Williams et al. 2019). This becomes particularly noteworthy when we consider the increasing levels of

online abuse, hatred and discrimination (Home Office, 2019). It is now easier than ever before to espouse a hateful message and reach audiences across the world in a matter of seconds. This is like a tsunami of hate, cyber-rippling across countries, causing offence, upset and pain.

This article will attempt to critically understand the motivational factors encouraging online hate speech. It will therefore draw on Goffman's (1959) ground-breaking work on self-presentation and then reconceptualise the model after taking into account the differences between online and offline communication. It will assess how these differences affect performances, and thus individuals' attitudes towards online derogation. The article explores how communication via new media technologies has blurred the boundaries between front and backstages. It also investigates the impact online hate exposure has on its audiences, and discusses the impact hate has on its victims. The article closes by examining the influence of the internet, and social media, and offers key suggestions for future research.

Online hate

Hate-speech is defined as spreading, inciting, or promoting hatred, violence and discrimination against an individual or group based on the their protected characteristics; which include 'race', ethnicity, religion, gender, sexual orientation, disability, among other social demarcations. If left unaddressed, hate speech can contribute towards acts of violence, and thus, hate crimes. The Home Office of the United Kingdom (UK) government (2019) reported that hate crimes have more than doubled in England and Wales since 2013, with the majority being racially motivated, accounting for 76% of all offences. And, between 2018-19, transgender attacks had increased by 37%, while anti-Semitic attacks had more than

doubled. Similarly, the F.B.I revealed that hate crimes had reached a 16 year high within the United States, with notable increases in attacks against Latino, transgender, and Sikh communities (Levin, 2019). Kilvington (2019) attempts to make sense of these increasing figures by contextualising the social and political climate across Europe and North America: 'When countries lurch through social and political unrest, racism and xenophobia are rarely far behind. Brexit and the changing political landscape in the US and Europe have seen debates around race, ethnicity and national belonging come to the fore'.

Therefore, we are living in tumultuous and hostile times as right-wing politicians have gained popularity and support, as observed within the United States (US), UK, Czech Republic, and France. Hateful and discriminatory offline behaviour is increasing and presents to us a very serious problem. But, to what extent are our online experiences shaping and reinforcing our offline behaviours, and vice-versa?

With the benefit of hindsight, it is possible to judge the *New Yorker's cartoon* from 1993, which proclaimed that 'On the Internet nobody knows that you are a dog' (Everett, 2009; Farrington et al. 2015; Steiner, 1993), as misguided. The cartoon postulated that 'one's identity was so hidden on the Web that opportunities would be widely open to all regardless of background characteristics that may have traditionally disadvantaged some people over others' (Hargittai, 2012: 224). Everett (2012: 165) adds that this example, among others, 'were symptomatic of [North America's] desire to imagine and construct colorblind or hyper-tolerant virtual communities and digital public spheres through the internet's text-driven digital environments during the late 1980s and early 1990s'. These examples present the internet as an 'idyllic, equal and even post-discriminatory world, one where everyone has a voice and the right to speak it' (Farrington et al. 2015: 43).

However, everyday offline experiences of oppression, disadvantage and prejudice are reflected in online experiences (boyd, 2011; Farrington et al. 2015; Nakamura, 2008;

Nakayama, 2017). Nakamura (2008) argues that the internet is an extraordinary example of a racialised medium while Nakayama (2017) notes that the culture of whiteness has been used to showcase, and reclaim, white supremacy within virtual spaces. People of colour have used online platforms to petition, protest and gather support, with Black Lives Matter being a case in point, when in contrast, whites are unable to receive attacks in the same way because the language for white oppression and discrimination does not transfer (Nakayama, 2017). For instance, the Labour politician, Diane Abbott, receives almost half of all abusive tweets that are sent to female politicians in the UK (Gayle, 2018). And, Farrington et al. (2015) discuss the case of former National Hockey League (NHL) player, Joel Ward, who, after scoring the winning goal in a cup final, received hundreds if not thousands of racist tweets which included death threats.

Not only are individuals targeted by perpetrators and trolls, but so too are communities and groups. Demos (2016) reported that worldwide, between March and June 2016, over 4 million Tweets were sent containing a word that could be considered anti-Islamic. Amnesty International (2018) reported that one in ten tweets that mention black women are considered abusive. The Inter-Parliamentary Coalition for Combating Anti-Semitism (ICCA) (2013) also noted that Facebook groups such as Kill a Jew Day, Kick a Ginger Day, Hitting Women, and Join if you Hate Homosexuals, had been created and generated online followers.

The Internet has allowed users a new platform to spread hate. Hate groups no longer have to communicate in isolation, hunt for new recruits, or distribute leaflets on foot. The Internet provides them instant access to new and existing followers, and it makes considerably easier to mobilise and spread hateful messages (Brown, 2009). Social media, in particular, has 'allowed old racial schemata to be broadcast in new social settings anonymously via smart phones and computers' (Cleland, 2014: 417). Hate speech is therefore no longer confined to offline spaces, as electronic communications devices are now being used to communicate and

spread hatred, allowing individuals and groups the opportunity to reach potentially larger audiences than ever before (Brown, 2009; Farrington et al. 2015; Kilvington and Price, 2019, 2017). As the ICCA (2013: 9) note, extremist and hate groups now 'host their own websites with impunity', using them to 'spread propaganda' and build communities of like-minded people.

Before exploring the motivations fuelling online hate speech, we must first discuss the Goffmanian theory of human communication.

Goffman: A theory of human interaction

The pioneering work of Erving Goffman (1922-1982) is central to our understanding of human behaviour and communication. Goffman (1959) was the first to employ a dramaturgical metaphor to help us make sense of the self within social encounters. Yet, Goffman was not the first to write about the self as Rogers' (1951) work attempted to theorise therapy sessions as patients attempted to *discover* the 'true self' in an attempt to interact more freely with others, while Jung's (1953) research distinguished between the unconscious self and the public mask, known as the persona. These ideas are, of course, influenced from Freud's archaeological model of the mind, which argues that the true self is hidden, buried beneath layers of defences, when engaging in the superficial practices of everyday social interaction. It was Goffman's theoretical model, however, that allowed us to further our understanding of how and why individual and group performances differ across social contexts.

The notion of the *façade* is particularly noteworthy. This is divided into two parts – the *stage* (setting, e.g. furniture, or in theatre speak, the props) and the *personal façade* (personal

front). The latter encompasses two distinct parts which include the *appearance* and the *manner*. The appearance refers to the person's external, or physical, self. Clothes, hair-style, make-up, among other visible characteristics, act as 'sign vehicles' which allow the observers to form judgements. These judgements are often influenced by drawing on previous experiences and histories with said 'sign vehicles'. The manner, on the other hand, relates to the person's behaviour, way of talking, and body language. When communicating on the stage, the personal façade enables the performer to both 'give' and 'give off' expressions. Expressions we give are intentional verbal symbols which we utilise to transmit information when attempting to generate a particular impression, while expressions given off are non-verbal and unintentional.

Performances are constructed with an awareness of societal and institutional expectations. In other words, interaction relationships are 'organised in particular by the use of shared resources and communication rules' (Serpa and Ferreira, 2018: 74). These rules are learned through repetition by drawing on factors that already exist within our repertoire. Masquerades are performed until the roles have been mastered. Teachers and pupils, nurses and patients, footballers and fans, all assume a role based on preconceived expectations of appropriate communication and behaviour. When an individual takes on an established role they are already likely to understand the rules and conventions due to prior exposure. Established roles, then, generate a 'collective representation' which is considered the norm.

Goffman (1959: 24) argues that individuals who perform in a certain social context exert a 'moral demand' upon their audience, encouraging them to respond in an expected or appropriate manner. As Durkheim (1926: 272) illustrates, then, 'we do not merely live but act; we compose and play our chosen character'. The portrait we paint is modified and tailored depending on the audience that we are interacting with. Put simply, individual performances are context specific as we decide which persona to adopt. Bullingham and

Vasconcelos (2013: 102) note that 'the self is merely the *mask* one chooses to wear in a given situation'. Goffman (1959) argues that individuals offer different performances within public (frontstage) and private (backstage) settings.

Like theatre, the front stage is public and includes an audience. Individuals therefore put on 'a show' and perform a desirable, expected, or anticipated image. Within the frontstage, actors are expected to follow the rules governing decorum and politeness in a bid not to offend. In the majority of cases, we display compassion and if this rule is broken, the actor may 'lose face'. Performers 'keep up appearances' as long as they remain within the frontstage, but 'terminate their performance when they leave it' (Goffman, 1959: 33).

Backstage interaction is antithetical to frontstage performances. Serpa and Ferreira (2018: 76) state that 'This is a restricted area, not in the public domain and without access to viewers'. In short, when the curtain comes down our public mask is removed. Goffman (1959: 129) argues that backstage language and behaviour includes 'profanity, open sexual remarks, elaborate griping, smoking, rough informal dress'. Moreover, it is within this private space where more honest, borderline and abhorrent views emanate (Farrington et al. 2015; Hylton, 2018, 2013; Kilvington and Price, 2019). Feagin and Picca (2007) and Hughey (2011) suggest that overtly racist communication has moved 'underground', shifting from frontstage to backstage spaces. Collective and derogatory terms, then, are employed to describe and label absent audiences which constructs and upholds an 'in-group-out-group split' (Goffman, 1959: 171). According to Goffman, derogation works to reaffirm the solidarity of the team, illustrating the mutual regard against the absent other. These private conversations contain 'dark secrets' which are kept hidden from the public through team solidarity. For an individual to 'fit in' to the team, they are expected to participate within these private and informal conversations while protecting their fellow team members. Bonds are developed and strengthened through 'team collusion'.

To what extent, though, can Goffman's theoretical model, notably around frontstage and backstage performances, be applied to the internet? In what ways does human interaction differ between online and offline worlds and how does this affect behaviour? And, is it possible to apply Goffman's existing model to virtual worlds in an attempt to critically understand the motivational factors encouraging online hate? These questions are of particular importance now as 'Hate speech has once more returned "over-ground", back centre stage, and is arguably more public than ever due to the advent of social media' (Kilvington and Price, 2019: 72).

Goffman, limitations, and online hate

Goffman's (1959) theory of human interaction and behaviour was exclusively based on the analytic interpretation of face-to-face situations where actors are physically co-present. Despite this, a number of scholars have applied Goffman's work to online spaces in a bid to critically understand online behaviour (Aspling, 2011; Bullingham and Vasconcelos, 2013; Merunkova and Slerka, 2019; Serpa and Ferreira, 2018). Fewer studies, however, have specifically employed Goffman's theoretical model to critically understand, and theoretically underpin, the motivations for online hate-speech (Hylton, 2018; Hylton and Lawrence, 2016; Hynes and Cook, 2013; Kilvington and Price, 2019). Although useful, it could be argued that Goffman's original model cannot simply be applied to online interaction when examining hate-speech because of the significant differences that exist between online and offline worlds, which influence and modify human behaviour.

This section will outline four key differences between online and offline communication. These differences are: anonymity, invisibility, dissociative imagination, and rapid response. It

is suggested that these factors, arguably made possible through online communication, encourage disinhibition which exacerbates online hate. Later, the article will develop and apply these factors within an updated and reworked Goffmanian model to help us critically understand the motivations for online hate. To date, work that applies Goffman's model to cyber-hate remains superficial and under developed as it fails to adequately demonstrate how performative stages are compromised, and blurred, when communicating via new media platforms.

Anonymity

Anonymity is widely regarded as a determining factor in online hate-speech. Online users have the ability to create fake accounts, or adopt pseudonyms which acts as a form of identity disguise which encourages disinhibition. That said, Santana's (2014) qualitative study on the topic of immigration within online forums found that while 65 percent of participants submitted hateful comments under the veil of anonymity, the remaining 35 percent posted uncivil comments using their real names. Whether anonymous or not, Suler (2004) outlines that the perception of anonymity encourages users to manipulate online expressions, view the Internet as a space where offline social norms do not apply, and take online interactions less seriously. In turn, anonymity, or perceived anonymity, works to 'embolden people to be more outrageous, obnoxious, or hateful in what they say than would be the case in real life' (Brown, 2017: 298-99). For Suler (2004: 322), anonymity affords users the 'opportunity to separate their actions online from their in-person lifestyle and identity' which results in feeling 'less vulnerable about self-disclosing and acting out'. Through the process of dissociation, users do not have to own their behaviour and instead, their online self becomes compartmentalised. In sum, disinhibition and deindividuation, facilitated through anonymity,

culminate in the polarisation of groups online. Arguably, these exacerbating factors of online hate are made possible through invisibility.

Invisibility

Belk (2014) writes that the self is now extended through digital expressions. But these digital expressions are disembodied, resulting in the image, sound and message essentially 'standing in' for, or replacing, the physical self. Goffman (1959) notes that when we enter another's presence, we make judgements based on 'sign vehicles'. Online, however, not only do we have less visible cues to drawn upon, but they are less trustworthy as digitally embodied identities might be fictitious or embellished.

Invisibility affects behaviour which can result in hate as aggressors are physically removed from their victims. As Farrington et al. (2015) state, 'If one cannot see others' physical expressions, one is less immediately aware of their dislike or distaste of our actions'. In short, online communication allows one to keep one's eyes averted which compromises the Goffmanian model as it is grounded in physical settings for human interaction. However, online communication is increasingly combining text, audio and visual interaction. Visual interaction is therefore no longer confined to offline communication. Hate can be espoused not only through text and images, but also through online live-streaming platforms. Although Farrington et al. (2015) and Suler (2004) have stated that perpetrators are unable to physically see their victims' reactions due to online invisibility, they can still be observed through representation, i.e. text, emoticons, videos, Gifs, etc. Despite being physically removed, perpetrators of online hate are still able to understand the consequences of their online expressions. But, significantly, perpetrators might never have to deal with the consequences of their actions as they might perform a 'hit-and-run' of hate, posting an abusive comment, and fleeing the scene of the crime, perhaps never to return (Suler, 2004). Because we can be

invisible, then, and potentially escape the 'crime scene' undetected, some users may perceive Internet communication as a game, a space without consequences, and where different rules apply.

Dissociative Imagination

For some, the internet feels like a 'make-believe dimension, separate and apart from the demands and responsibilities of the real world' (Suler, 2004: 323). Users are therefore able to create personas or embellish aspects of their identities because they disassociate online fiction from offline fact. Suler (2004: 323 emphasis added) adds that when users 'turn off their computer and return to their daily routine, they believe they can *leave behind that game and their game identity*'. Now, because technology, for example, has become part of the extended self, and embedded within every day practices, we consistently navigate or switch between online (perceived as not-real) and offline (real) interactions (Battin, 2017).

The act of trolling is thus important to consider. A troll is an internet user who attempts to intentionally disrupt the online community or individual by causing upset, offense and trouble. Berghel and Berleant (2018) put forward a taxonomy of trolling which includes provocation, social-engineering, sport, rehearsal, insult, and satire trolling. A number of these categories relate to the concept of dissociative imagination, especially sport, rehearsal and insult trolling as this behaviour is considered a game which leads to self-gratification. Cheng et al. (2017) illustrate that trolling, like laughter, can be contagious, and that ordinary people, placed within certain conditions, can become trolls. They add that a growing number of online users are beginning to engage in trolling behaviour within online news forums, finding that one fifth of comments uploaded to CNN.com are removed by moderators for violating community guidelines. Nakayama (2017) continues, suggesting that people are encouraged to

perform in excessive ways online in an attempt to 'out-do' the last comment or post by being more outrageous in an attempt to gain more attention and reactions.

Sanctions have the power to shift the perception of online behaviour from a make-believe dimension to a space which has 'real-world' consequences for the aggressors. If virtual worlds are perceived as separate, it is likely to lower our self-awareness. Coupling dissociative imagination with the desire for instantaneous communication on social media sites, such as Twitter, leads to posts being disseminated without appropriate consideration of the communicative content. The reactionary nature of social media perhaps influences online hate (Brown, 2017; Farrington et al. 2015).

Rapid Response

Many scholars have argued that the Internet has taken over our daily lives, with social media being at the forefront (Farrington et al. 2015; Nakamura, 2008). Our smart phone, tablet or laptop are usually at arm's length, or within the palm of our hand, when consuming our favourite television programme or watching a football match or concert. In many cases, these devices are actually used to stream said content. For McGillivray and McLaughlin (2019: 33), 'The media landscape has changed and websites, online content, and social media are acting as "second screens" to the primary broadcast via television and are being used simultaneously by fans'.

The reliance and desire to communicate online has facilitated an 'addictive call and response' feeling (Manghani, 2009). Serpa and Ferreira (2018) state that social media relies on immediate reactions and that if one has less time to think, then it enhances one's chances of posting something hateful. For Brown (2017: 304), 'the Internet encourages forms of hate speech that are spontaneous in the sense of being instant responses, gut reactions, unconsidered judgements, off-the-cuff remarks, unfiltered commentary, and first thoughts'.

Trigger events are therefore worth considering as reactive social media posts tend to be driven by emotion, almost like a 'knee-jerk reaction' (Suler, 2004). Triggers in sport, such as a goal being scored or a penalty kick being missed, have resulted in the rapidity of online racism (Busby, 2019; Burrows, 2019; Farrington et al. 2015). Stephan and Stephan's (2000) work on realistic and symbolic threats helps us critically understand responses to trigger events. While realistic threats refer to resources such as jobs and welfare, symbolic threats relate to perceived challenges towards ideologies and values. Realistic threats tend to challenge individuals' livelihood, while symbolic threats are paradoxical to the 'ingroups' collective belief system. In turn, triggers can lead to online posts which showcase automatic prejudice and instant stereotyping, while derogatory language is used without awareness and consideration.

Rapid response exacerbates online hate, but not all social media platforms encourage rapidity, and this is a vital distinction. Twitter, Snapchat, and WhatsApp are built on spontaneity whereas LinkedIn, blogs and online forums encourage more planned, considered, non-spontaneous forms of speech whereby hateful and borderline views are carefully constructed to entice targeted audiences, based on collective worldviews, senses of grievance, and so on. Bluic et al. (2018) state that groups and individuals perform hate online for different reasons or to achieve different outcomes. They argue that those belonging to a group tend to promote and defend hateful ideologies while individuals simply perform hate to hurt outgroups in response to trigger events. It is more likely, then, that individuals belonging to racist or xenophobic groups will perhaps provide more considered responses in an attempt to justify or cultivate their ideology, in contrast to individuals who display one-off or intermittent outbursts of hate based on social, cultural or political triggers. The work of Richey et al. (2018) illustrate that Twitter users are aware that spontaneity is encouraged, and integral to the sites design. Yet, the participants noted that although prolonged reflection was

not encouraged, it should be embraced as it allows users to take control, thus avoiding posting morally and socially damaging content.

This section has critically explored the differences between offline and online interaction and behaviour. Offline, we can be anonymous by wearing a mask, a balaclava, or a hood. Offline, we can be invisible by writing graffiti on a public wall or by hiding while shouting derogatory slurs at passer-by's. However, offline spaces are not perceived to be 'make-believe dimensions' as social norms and sanctions generally govern behaviour within public spaces. Offline, high-prejudiced people may suppress instant stereotypes in public settings while in contrast, some online platforms encourage first thoughts which exacerbates online hate. Although the features offered to individuals online are not entirely unique to Internet communication (Brown, 2017), these differences are still significant enough to seriously contest and challenge Goffman's theoretical model of communication.

We cannot simply apply Goffman's model to virtual communication when attempting to understand the reasons for uploading online hate. Goffman's work must be comprehensively rethought, reshaped and remodelled as virtual frontstages and virtual backstages have blurred. We now live in a world where it is arguably considered more acceptable to abuse another person online, rather than offline (Nakayama, 2017). The internet has provided a platform where it is easier, quicker, and cheaper to spread hate, leading to group polarisation and mobilisation. The following section explores virtual frontstages and virtual backstages.

Exploring Virtual Performances

Goffman (1959) postulates that while performances are in process, 'audience segregation' occurs; meaning that actors play different parts in other settings. Thus, the actor adopts

different personas depending on the expectations of the audience. In conjunction with Goffman's theoretical framework, then, Bullingham and Vasconcelos (2013: 102) note that 'online environments provide users with the potential to perform and present different identities'. Identity is constructed online, as we choose to project given identities within appropriate social settings. Our performances are tailored and constructed around the rules, conventions and norms that exist within different virtual environments. By applying Goffman's theory of front and backstage performances, and impressions given and given off, to Internet communication, it allows us to conceptually understand individual communication, interaction, and behaviour across social media platforms. Put simply, we now have virtual frontstages and virtual backstages (Aspling, 2011).

Virtual frontstages are regarded as open spaces as audiences are able to browse user profiles and make judgements about their character and behaviour (Moore et al. 2017). Due to the perceived nature of Facebook, Instagram, Twitter and LinkedIn, individuals modify their personas and behaviour based on their perceptions of that space. Merunkova and Slerka (2019: 271) note that Facebook users 'build their image using profile and cover photos, shared posts, their interests and also photos where they tag friends. They post only the inoffensive and desirable ones on their profile'. Goffman's (1959) work on impression management is particularly noteworthy here as Wallace et al. (2014) illustrate that Facebook users tend to consciously 'like' other users' posts to provide a positive, supportive and friendly impression of the self. Belk (2014) adds that the practice of 'liking' and providing flattering comments is common because recipients of such affirmations tend to reciprocate for their friends as well. As a result, this online practice appears spontaneous whereas it could be regarded as egotistical self-love. Furthermore, Merunkova and Slerka (2019) state that users delete posts which acquire no feedback. The absence of reaction and interaction illustrates that the post is bad, or boring, and gives off a damaging impression to the audience

where the actor loses face. Virtual frontstage performances, then, can be carefully constructed and shaped by individuals as we have the power to decide what information is public and private.

The virtual backstage offers a more personal, relaxed and honest performance (Merunkova and Slerka, 2019; Serpa and Ferreira, 2018). This communication, considered safe and secure, takes place via closed, direct and private messaging platforms including Facebook, WhatsApp and SMS messaging. Merunkova and Slerka (2019: 268) state that Facebook users see the border between virtual frontstages and virtual backstages relatively clearly, adding that 'users decide which information to post "publicly" on their profile … and which information they intend to share only with selected persons via private messages'. Like offline backstage performances, bonds and solidarity are formed within virtual backstage spaces meaning that these private interactions remain concealed and protected through 'team collusion' (Goffman, 1959). Hylton and Lawrence (2016) and Hynes and Cook (2013) postulate that hidden virtual backstages are being used to perform derogation and thus team solidarity helps prevent against unwanted exposés. For Cain (2012: 669), 'The two regions have a symbiotic relationship in that activities in the backstage allow workers to maintain appropriate behaviors during the front stage, while activities provide fodder for discussions and activities in the back region'. Hylton and Lawrence (2016) highlight instances where public figures have been exposed for committing acts of derogation within virtual backstages which have resulted in consequences for the perpetrators. These exposés have the ability to undermine the carefully constructed personas that have been created within frontstage and virtual frontstage spaces.

Although individuals often use virtual frontstage spaces to purport an idealised version of the self (Merunkova and Slerka, 2019; Miguel and Medina, 2010; Moore et al. 2017; Serpa and Ferreira, 2018), we must consider the influence that anonymity, invisibility, dissociative

imagination and spontaneity have upon performances (Brown, 2017; Farrington et al. 2015; Kilvington and Price, 2019, 2017; Suler, 2004). Online derogation is not just reserved for virtual backstages; it has saturated virtual frontstages as publicly visible online hate speech continues to increase (Cheng et al. 2017; Kilvington and Price, 2019, 2017). This seriously compromises Goffman's (1959) original model of self-presentation as virtual frontstage performances are not always idealised. The nature of communication, and the conditions in which we compose our communication, therefore impacts upon our online social practices and behaviours. While Goffman (1959) suggested that frontstage and backstage performances were separate and distinct, it could be argued that virtual stages have blurred which exacerbates online hate speech.

The Virtual Stages of Hate: What Motivates Cyberhate?

New technology is changing our world, and these differences are affecting human communication, interaction, and behaviour. Arguably, all virtual communication, whether it is posted in a virtual frontstage or a virtual backstage, is created or composed within a space that simulates backstage feelings of privacy, safety and security. Problematically, 'The line between private and public is blurred in the context of social media' because virtual frontstages often feel private and personalised despite being public (Merunkova and Slerka, 2019: 271). It could be suggested that as the communicative content is being composed, the output of the message, and thus the intended audience, is not being fully acknowledged. It is this 'backstage mimicry', encouraged by factors including anonymity and invisibility, that has arguably led to an increase in online hate. Figure 1.1 offers a revised Goffmanian model and attempts to illustrate how online performances are affected through feelings of

disinhibition. These differences enhance feelings of courage and freedom (Brown, 2017; Farrington et al. 2015; Keum and Miller, 2018; Kilvington and Price, 2019, 2017; Suler, 2004) and blur virtual frontstages and backstages.

<INSERT FIGURE 1.1 – The Virtual Stages of Hate>

If this behaviour continues, it desensitises us to it (ICCA, 2013). Hawdon et al's. (2017) study reported that an average of 43 per cent of participants, aged between 15 to 30 years old, in the US, UK, Germany and Finland had encountered online hate material. Most of the online hate material was experienced on social networks sites, such as Twitter and Facebook. For Williams et al. (2019: 6), 'Far right and popular right-wing activity on social media, unhindered for decades due to free-speech protections, has shaped the perception of many users regarding what language is acceptable online'. In turn, our online experiences are helping shape our ideologies, belief systems and behaviours.

The impact of filter bubbles is noteworthy. It has been suggested that the more we contribute to these bubbles, by liking posts and following other like-minded individuals and groups, algorithms help neaten the bubble. Because ranking algorithms work to filter out posts, 'echo chambers' are created on social media whereby increasingly extreme viewpoints are being consumed (Sunstein, 2007). These 'echo chambers' are becoming breeding grounds for division and radicalisation (Williams et al. 2019). If an individual spends weeks, months or years living inside a filter bubble where racist, sexist or homophobic views are posted with impunity, desensitisation may occur. Essentially, filter bubbles are helping blur those virtual stages and equipping some people with the confidence to post discriminatory material online.

Our experience of news media through social media platforms is also relevant when exploring filter bubbles because it has been suggested that media organizations create

agendas which are then discussed by online audiences, as opposed to audiences setting their own agendas (McCombs, 2014; Williams et al. 2019). Pew Research Centre (2018) reported that online sources, including social media, are now more popularly consumed than traditional press outlets for news in the US while 'two-thirds of UK adults, and eight in ten 16 to 24 year olds now use the Internet as their main source of news' (Williams et al. 2019: 7). As a result, algorithms are personalising our social media experiences of news but rather than being challenged on belief systems, users' viewpoints are being reinforced through the filter bubble effect (Bruns, 2019). This can be particularly damaging as persistent negative framing of communities and groups can influence attitudes and behaviours (Goffman, 1974). Tuchman's (1978) work on symbolic annihilation is crucial as symbolic representations and portrayals are a form of power, and if some groups or communities are omitted from the bubble, it illustrates and cements their devalued status.

The younger generation are now growing up online, living in a hostile, unfiltered, and unregulated world which is helping shape their ideologies through social media consumption (Merunkova and Slerka, 2019). Lind (2019: 1) states that 'Worldwide, the average internet user is on social media more than five hours per day' meaning that the filter bubbles' edges become firmer until they are almost impenetrable. Belk (2014) argues that Internet users are invested in social media as technological devices become extensions of the self. We are not, however, attached to the physical devices for communication; we are attached to what they allow us to do. They allow us to interact, learn, engage, comment, share, respond, like, compliment, and discuss. Conversely, they also facilitate the spreading of fake news, misinformation, violence, hate, discrimination, bullying, and abuse. Belk (2014) asks, will this online hate leak into the offline world?

In response, research illustrates that where hate speech rises online, it simultaneously rises offline (Williams et al. 2019). For example, the United Nations (UN) stated that Facebook

played a ‘determining role’ in stirring up hatred against the Rohingya Muslims in Myanmar (Smith, 2018), while social media was used to fuel the civil war in South Sudan (McCarthy, 2017). Although trigger events such as immigration, civil unrest, election results and terror attacks are responsible for increasing racial and religious tensions, Williams et al’s. (2019) research shows that online hate speech is a process as social media users continually discuss and debate divisive agendas set by media organisations. Williams et al’s. (2019: 15) superb work indicates ‘a strong link between hateful Twitter posts and offline racially and religiously aggravated crimes in London’, adding that a rise of 1,000 hate tweets would see a ‘4 per cent increase in racially or religiously aggravated harassment in a given month’ (Williams et al. 2019: 19). Hence, it is essential that research into online communities, behaviours and practices is undertaken. It is vital that we understand the factors driving online hate, and what the continued exposure to hate and misinformation is having on contemporary audiences.

In summary, Goffman (1959: 45) said that ‘the world, in truth, is a wedding’. It is a space where audiences wear their happiest masks. Honesty is supressed. We smile, laugh and adopt the expected character for the duration of the performance. As we navigate frontstage spaces we constantly attempt to be the best versions of ourselves. Yet, is this happening online? Many virtual spaces have become toxic, littered with nasty insults, bullying, abuse, discrimination and hate.

Goffman’s remodelled work allows us to understand the motivations for posting online hate. The key difference is that online communication is composed within a blurred space which mimics the backstage which affects the communicative act. As a result, this affects performances and behaviours. The concept of ‘backstage mimicry’ offers us clear insights into contemporary audiences’ concealed thoughts and illustrates the importance of research

into online communities; it shows that we still have a long way to go in the fight against online hate, violence and discrimination.

Conclusion

This article has attempted to build a theoretical model, inspired by Goffman's (1959) work, to help understand the factors encouraging online hate speech. Once we understand this behaviour we can better attempt to challenge it and plot its demise. It has illustrated that the differences between online and offline communication affect performances which impacts behaviours. It suggests that boundaries between frontstages and backstages have blurred in regards to online communication. As a result, researchers are now able to observe contemporary audiences' attitudes and ideologies which were once reserved for the backstage.

We must also consider the impact that online exposure to hatred and discrimination is having on internet users. Kozinets (2002) argues that virtual social groups influence the members who participate in them which affects their behaviour both online and offline. Algorithms, which cause filter bubbles, are problematic in that they reinforce attitudes rather than challenge them (Bruns, 2019). As Williams et al. (2019) note, social media profiles have become echo chambers which are becoming breeding grounds for polarisation and radicalisation. It could be suggested that users' experience of social media is affecting how audiences understand the world, and their place within it. With greater exposure to certain ideologies, such views arguably become normalised. Therefore, through the features which the internet offers, it instils a level confidence in our beliefs and in some cases, encourages individuals to post contentious and inflammatory content. For some people, they are

beginning to turn online hate speech into real-world offline hate speech and hate crimes (Williams et al. 2019). Of course, social media is not solely responsible for the increase of hate speech (Kilvington, 2019) as perpetrators' empathy levels (Brown, 2017), moods (Farrington et al. 2015), and the contexts of conversations (Cheng et al. 2017) must be considered.

Online hate is increasing and evolving which emphasises the importance and timeliness of this work. Not only are victims likely to experience abuse more online than offline, they are also more likely to experience hateful events or acts repeatedly and in different formats because expressions of hate can now be transmitted in multiple forms including texts, photos and GIFs (Keum and Miller, 2018). Research on the impact of online victimisation is paramount. Surprisingly, however, Bluic et al. (2018) state that only four academic studies between 2005 to 2015 explicitly focused on the effects of cyber-racism. Nonetheless, for Brown (2017: 307), online abuse might be more damaging than offline abuse because of 'the volume of abuse facilitated by online communication. Moreover, it might be that online hate speech has especially harmful effects because it is done in front of larger audiences, thus ramping up the public shame element'. Overt racism, for example, is associated with depression, low self-esteem and anxiety (Feagin and Elias, 2013; Tynes et al. 2008). These consequences also correlate with online abuse as Bluic et al. (2018: 84) state that 'cyber-racism has effects which are of similar or higher levels to those of discrimination experienced offline'. Put simply, empirical research focusing specifically on the psychological impact that online hate, bullying and discrimination has on its victims is essential.

In sum, this work has outlined the motivational factors encouraging online hate speech by reconceptualising Goffman's (1959) influential work. Future research should attempt to empirically test *The Virtual Stages of Hate* to further understand online hate. Moreover, further research should be conducted into algorithms and filter bubbles to critically examine

the ways in which they fragment and polarise audiences. Crucially, it is fundamental that victims' experiences of online abuse, hate and discrimination are empirically captured, explored and understood. Once we understand the impact of this behaviour, we are in a stronger position to challenge it. The virtual world, in truth, is certainly not a wedding. But with greater research, education and action, it is hoped that the virtual stages of hate can become the virtual stages of inclusion and respect.

References

Amnesty International UK. (2018) Women abused on Twitter every 30 seconds - new study. Available at: https://www.amnesty.org.uk/press-releases/women-abused-twitter-every-30-seconds-new-study (accessed 11 February 2020).

Aspling, F. (2011) *The Private and the Public in online Presentations of the Self: A Critical Development of Goffman's Dramaturgical Perspective*. MA. Stockholm University.

Battin, J. (2017) *Mobile Media Technologies and Poiesis: Rediscovering how we use technology to cultivate meaning in a nihilistic world*. Cham: Palgrave McMillan.

Belk, R. (2014) 'Digital consumption and he extended self', *Journal of Marketing Management* 30(11-12): 1101-1118.

Berghel, H. and Berleant, D. (2018) The online trolling ecosystem. *Aftershock*. Available at: http://www.berghel.net/col-edit/aftershock/aug-18/aftershock_8-18.pdf (accessed 11 February 2020).

Bluic, A.M., Faulkner, N., Jakubowicz, A. and McGarty, C. (2018) 'Online networks of racial hate: A systematic review of 10 years of research on cyber-racism', *Computers in Human Behavior* 87: 75-86.

boyd, D. (2011) 'White flight in networked publics? How race and class shaped American teen engagement with MySpace and Facebook', pp.203-222 in L. Nakamura and P. Chow-White (eds) *Race After the Internet*. London: Routledge.

Brown, C. (2009) 'WWW.HATE.COM: White supremacist discourse on the Internet and the construction of whiteness ideology', *The Howard Journal of Communications* 20(2): 189-208.

Brown, A. (2017) 'What is so special about online (as compared to offline) hate-speech?' *Ethnicities* 18(3): 297-326.

Bruns, A. (2019) *Are Filter Bubbles Real?* London: Polity Press.

Bullingham, L. and Vasconcelos, A.C. (2013) 'The presentation of self in the online world: Goffman and the study of online identities', *Journal of Information Sciences* 39(1): 101-112.

Burrows, B. (2019) Manchester United 'disgusted' by racist abuse aimed at midfielder following Wolves penalty miss. *The Independent*, 20 August. Available at: https://www.independent.co.uk/sport/football/premier-league/manchester-united-wolves-result-paul-pogba-penalty-racist-abuse-twitter-club-statement-a9071076.html (accessed 11 February 2020).

Busby, M. (2019) Manchester United's Marcus Rashford target of racist abuse on Twitter. *The Guardian*, 24 August. Available at: https://www.theguardian.com/football/2019/aug/24/manchester-uniteds-marcus-rashford-target-of-racist-abuse-on-twitter (accessed 11 February 2020).

Cain, C.L. (2012) 'Integrating dark humor and compassion', *Journal of Contemporary Ethnography* 41(6): 668-694.

Cheng, J., Bernstein, M., Danescu-Niculescu-Mizil., C. and Leskovec, J. (2017) 'Anyone can become a troll: Causes of trolling behavior in online discussions', *ACM Conference on Computer Supported Cooperative Work and Social Computing*. Available at: https://dl.acm.org/doi/proceedings/10.1145/2998181

Cleland, J. (2014) 'Racism, football fans, and online message board: How social media has added a new dimension to racist discourse in English football', *Journal of Sport and Social Issues* 38(5): 415-431.

Demos. (2016) Islamophobia on Twitter, 18 August. Available at: https://demos.co.uk/project/islamophobia-on-twitter/ (accessed 11 February 2020).

Durkheim, E. (1926) *The Elementary Forms of Religious life*. London: Allen & Unwin.

Everett, A. (2009) *Digital Diaspora: A Race for Cyberspace*. New York: SUNY Press.

Feagin, J. and Elias, S. (2013) 'Symposium on rethinking Racial Formation Theory: A Systemic Racism critique', *Ethnic and Racial Studies* 36(6): 931-960.

Feagin, J. and Picca, LH. (2007) *Two-Faced Racism: Whites in the Backstage and Frontstage*. New York: Routledge.

Farrington, N., Hall, L., Kilvington, D., Price, J. and Saeed, A. (2015) *Sport, Racism and Social Media*. London: Routledge.

Gayle, D. (2018) Diane Abbott: Twitter has 'put racists into overdrive.' *The Guardian*, 18 December. Available at: https://www.theguardian.com/politics/2018/dec/18/diane-abbott-calls-for-twitter-to-clamp-down-on-hate-speech (accessed 11 February 2020).

Goffman, E. ([1959] 1990) *The Presentation of Self in Everyday Life*. London: Penguin Books.

Hargiatti, E. (2012) 'Open doors, closed spaces? Differentiated adoption of social network sites by user background', pp.223-245 in L. Nakamura and P. Chow-White (eds) *Race After the Internet*. London: Routledge.

Hawdon, J. Oksanen, A. and Rasanen, P. (2017) 'Exposure to online hate in four nations: A cross-national consideration', *Deviant Behavior* 38: 247-67.

Hine, C. (2012) *The Internet: Understanding Qualitative research*. Oxford: Oxford University Press.

Home Office. (2019) Hate Crime, England and Wales, 2018-19. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/839172/hate-crime-1819-hosb2419.pdf (accessed 10 February 2020).

Hughey, M. (2011) 'Backstage discourses and the reproduction of white masculinities', *Sociological Quarterly* 52(1): 132-153.

Hylton, K. (2018) *Contesting 'Race' and Sport: Shaming the Colour Line.* London: Routledge.

Hylton, K. (2013) 'Dispositions to 'race' and racism on the internet: online reactions to a 'racist' comment against Tiger Woods', *CERS Working Paper*. University of Leeds.

Hylton, K. and Lawrence, S. (2016) ' "For your ears only!" Donald Sterling and backstage racism in sport', *Ethnic and Racial Studies* 39(15): 2740-2757.

Inter-Parliamentary Coalition for Combating Anti-Semitism (ICCA). (2013) Report by the task force on internet hate, 29 May. Available at: http://www.report-it.org.uk/files/icca_task_force_on_internet_hate_report_may_29_2013_final.pdf (accessed 11 February 2020).

Hynes, D. and Cook, A.M. (2013) 'Online belongings: Female fan experiences in online soccer forums', pp.97-110 in B. Hutchins and D. Rowe (eds) *Digital Media Sport: Technology, Power and Culture in the Network Society*. New York: Routledge.

Jung, CG. (1953) *Two Essays on Analytical Psychology*. London: Routledge & Kegan Paul.

Keum, B.T. and Miller, M.J. (2018) 'Racism on the Internet: Conceptualization and recommendations for research', *Psychology of Violence* 8(6): 782-791.

Kilvington, D. (2019) 'Racist abuse at football games is increasing, Home Office says – but the sport's race problem goes much deeper', *The Conversation*, 9 October. Available at: https://theconversation.com/racist-abuse-at-football-games-is-increasing-home-office-says-but-the-sports-race-problem-goes-much-deeper-124467 (accessed 10 February 2020).

Kilvington, D. and Price, J. (2017) 'Tackling social media abuse? Critically assessing English football's response to online racism', *Communication & Sport*. Available online: http://eprints.leedsbeckett.ac.uk/view/creators/Kilvington=3ADJ=3A=3A.html

Kilvington, D. and Price, J. (2019) 'From backstage to frontstage: Exploring football and the growing problem of online abuse', pp.69-85 in S. Lawrence and G. Crawford (eds) *Digital Football Cultures: Fandom, Identities and Resistance*. New York: Routledge.

Kozinets, R.V. (2002) 'The field behind the screen: Using netnography for marketing research in online communities', *Journal of Marketing Research* 39: 61-72.

Levin, S. (2019) 'Violent hate crimes in US reach highest levels in 16 years, FBI reports', *The Guardian*, 12 November. Available at: https://www.theguardian.com/society/2019/nov/12/hate-crimes-2018-latinos-transgender-fbi (accessed 10 February 2020).

Lind, R.A. (2019) 'Laying a Foundation for Studying Race, Gender, Class, and the Media', pp.1-10 in RA. Lind (ed) *Race/Gender/Class/Media*. New York: Routledge.

Manghani, S. (2009) 'Love messaging: Mobile phone txting seen through the lens of Tanka poetry', *Theory, Culture & Society* 26(2-3): 209-232.

McCarthy, J. (2017) 'Fake news in South Sudan could lead to genocide', *Global Citizen Online*, 18 January. Available at: https://www.globalcitizen.org/en/content/fake-news-in-south-sudan-could-lead-to-genocide/ (accessed 12 February 2020).

McCombs, M. (2014) *Setting the Agenda: The Mass Media and Public Opinion* (2nd edition). Cambridge: Polity.

McGillivray, D. and McLaughlin, E. (2019) 'Transnational digital fandom: club media, place, and (networked) space', pp.30-46 in S. Lawrence and G. Crawford (eds) *Digital Football Cultures: Fandom, Identities and Resistance*. New York: Routledge.

Merunkova, L. and Slerka, J. (2019) 'Goffman's theory as a framework for analysis of self presentation on online social networks', *Masaryk University Journal of law and Technology* 13(2): 243-276.

Miguel, C. and Medina, P. (2010) 'The transformation of identity and privacy through online social networks (the CouchSurfing case)', *Social Media, Networks and Life. McLuhan Galaxy Conference*, 331-342.

Moore, C., Barbour, K. and Lee, K. (2017) 'Five dimensions of online persona', *Persona Studies* 3(1): 1-11.

Nakamura, L. (2008) *Digitizing race: Visual culture on the Internet.* Minnesota: Minnesota Press.

Nakayama, T. (2017) 'What's next for whiteness and the Internet', *Critical Studies in Media Communication* 34(1): 68-72.

Richey, M., Gonibeed, A. and Ravishankar, M.N. (2018) 'The perils and promises of self-disclosure on social media', *Information Systems Frontiers* 20(3): 425-437.

Rogers, C.R. (1951) *Client-Centered Therapy*. Boston: Houghton Mifflin

Santana, A.D. (2014) 'Virtuous or vitriolic: The effect of anonymity on civility in online newspaper reader comment boards', *Journalism Practice* 8: 18-33.

Serpa, S. and Ferreira, C.M. (2018) 'Goffman's backstage revisited: Conceptual relevance in contemporary social interactions', *International Journal of Social Science Studies* 6(10): 74-80.

Stephan, W.G. and Stephan, C.W. (2000) 'An integrated threat theory of prejudice', pp.23-45 in S. Oskamp (ed) *Reducing Prejudice and Discrimination*. Mahwah, NJ: Lawrence Erlbaum Associates.

Tuchman, G. (1978) 'The symbolic annihilation of women by the mass media', pp.5-38 in G. Tuchman., A.K. Daniels. and K.W. Benet (eds) *Health and Home: Images of Women in the Mass Media*. New York: Oxford University Press.

Tynes, B.M., Giang, M.T., Williams, D.R. and Thompson, G.N. (2008) 'Online racial discrimination and psychological adjustment among adolescents', *Journal of Adolescent Health* 43: 565-569.

Smith, N. (2018) 'UN says that Facebook helped fuel Rohingya ethnic cleansing', *The Telegraph*, 13 March. Available at: https://www.telegraph.co.uk/news/2018/03/13/un-says-facebook-fuelled-rohingya-ethnic-cleansing/ (accessed 12 February 2020).

Steiner, P. (1993) 'On the internet nobody knows you're a dog', *New Yorker* 69:61.

Suler, J. (2004) 'The online disinhibition effect', *Cyber Psychology & Behavior* 7(3): 321–326.

Sunstein, C. (2007) *Republic.com 2.0*. Princeton: Princeton University Press.

Wallace, E., Buil, I., De Chernatony. L. and Hogan, M. (2014) 'Who "likes" you… and why? A typology of Facebook fans from 'fan'-atics and self expressiveness to utilitarians and authentics', *Journal of Advertising Research* 54(1): 92-109.

West, R. and Thakore, B. (2013) 'Racial exclusion in the online world', *Future Internet*, 5(2): 251-267.

Williams, M.L., Burnap, P., Javed, A., Liu, H. and Ozalp, S. (2019) 'Hate in the machine: Anti-black and anti-Muslim social media posts as predictors of offline racially and religiously aggravated crime', *Crime and Justice Studies*. Oxford: Oxford University Press.