Hate Speech Crimes in Cyberspace: A Criminological/ Communicological Investigation of Social Media in South Africa – Understanding the Cyber-mind

Ian P. Saunderson
Cornelis Roeloefse[1]
Christopher Gumbi

Abstract

This contribution examines hate speech in the cyberspace in South Africa from both a Criminological and Communication Studies perspective. From the field of Criminology, legislation and interpretation of unlawfulness, intent and incitement of hate speech are examined. From the field of Communication Studies conceptualisations surrounding fragmented identities, Goffman's dramaturgy and the Spiral of Silence theory are used to analyse hate speech in the cyberspace through a cross-disciplinary interpretation. A qualitative, thematically orientated methodology is employed which allows for, not only understanding of online communication and meta-theoretical conceptualisations of the cyber-mind, but classification in terms of legal aspects surrounding hate speech. Data were collected from social media postings nationally over the past year, utilising a convenience sample; and classified into two cohorts, (1) postings from the ordinary person; and (2) postings from politicians. Findings are interpreted against legal conceptualisations of hate speech whilst utilising the above communication theories for the provision of a meta-interdisciplinary-perspective. Findings highlight the difference between threats posted on social media by ordinary people and is compared to instruction posted by

[1] Prof C. Roelofse, late of the University of Zululand, 2019. This is a posthumous submission. Prof Roelofse was a colleague of the other two authors before moving to the University of Zululand in January 2019.

Print ISSN 1023-1757; Electronic ISSN: 2519-5476; DOI https://doi.org/10.29086/2519-5476/2019/sp24.2a13

politicians. The ramifications of these two clearly distinguishable acts of communication are discussed first from a legal perspective and then contrasted against communication theoretical orientations of fragmented identities, Goffman's dramaturgy and the spiral of silence in attempt to understand the cyber-mind.

Keywords: hate speech, social media, incitement, genocide, communication and criminological theory

Introduction

The use of social media has been growing exponentially. The communication potential of social media on Facebook alone demonstrated 2.2 billion monthly active users in the last quarter of 2017 (Statista.com 2018:1). At the same time, the micro-blogging service of Twitter had 330 million monthly active users on average. With these vast numbers of people online, just mentioning two of the many available services, communication with a worldwide audience is at the fingertips of its users. The share facility makes it possible that one post or tweet can reach millions of users in a short space of time.

The theoretical orientation of this article is based on crimino/legal conceptualisations of hate speech, unlawfulness, intent, incitement and instructtion. From a Communication Studies perspective Goffman's dramaturgical theory (Goffman 1959), Identity management theory (Bapista 2003) and the Spiral of Silence theory (Noelle-Neumann 1974) are utilised. Hate speech manifests online, posted by individuals who have a certain image to maintain (Bapista 2003), tries to manifest a certain identity in society online (which may differ even substantially from the identity of the person in real life) and are posturing, performing and playing to political, religious or some other overture. Within this context Noelle-Neumanns' (1974) Spiral of Silence Theory is re-interpreted for the purpose of social media analysis. The strength of Noelle-Neumann's theory lies in the fact that it allows for interpretation of communication from the masses who would under normal circumstances remain silent and follow the opinions leaders in society. During times of political uncertainty, the masses let their voices be heard, as is clearly the case in this analysis.

Whilst the criminological investigation is a clear indicator of the extent to which hate speech legislation is contravened on social media, the communi-

cological investigation allows for a broader theoretical interpretation measuring communicological concepts within society. The choice of methodology, thematic analysis, allows for the exposure of agentic tools and how they are interplayed between politicians and the ordinary social media user, as a re-interpretation of the reasonable/ordinary person concept.

A Criminological Perspective on Hate Speech

This section first examines the supreme and overarching law in South Africa, the Constitution, Act 108 of 1996. It specifically focuses on section 16, which deals with freedom of expression. This is followed by a brief discussion of genocidal hate speech in international law, followed by an examination of South African criminal and statutory law in relation to hate speech. Lastly this section examines the intent of hate speech under the auspices of threat and instruction, against the backdrop of wrongfulness, as it manifests on social media.

The South African Constitution

Freedom of speech is entrenched in constitutional democracies. This ensures robust debate, brings challenging issues into the public domain and engages government and challenges its policies, actions and omissions. Section 16 of the *Constitution of the Republic of South Africa* (1997) deals with freedom of expression in sub-section 1, but also addresses the limitations of the freedom in sub-section 2.

Section 16 of the Bill of Rights specifically delimits freedom of expression to:

a. Press and other media;
b. Receive or impart information and/ or ideas;
c. of artistic creativity; and
d. academic freedom and freedom of scientific research.

The South African Constitution expressly list categories of these freedoms that are forbidden expressions as listed in sub-section 2, namely:

a. Propaganda for war;

b. Incitement of imminent violence; or
c. Advocacy of hatred that is based on race, ethnicity gender or religion and that constitutes incitement to cause harm.

The criminological perspective of the article aims at analysing data collected form social and written media and to apply the freedoms granted under sub-section 16(1) against the limitations of sub-section (2). The limitations do not pose a particular problem in interpretation apart from 'propaganda for war'. The question is whether this includes civil war? Thefreedictionary.com (2018:1), inter alia defines war as, '… conflict between nations or factions within a nation..'.. The English Dictionary (2018:1) confirms the definition by stating that war is 'open armed conflict between countries or between factions within the same country'. In this article, propaganda for war is deemed as incitement of violence between factions within a nation. From these definitions the premise is derived that freedom of expression explicitly forbids propagating war between factions in South Africa.

International Law

Boggenpoel (2013) writes that:

> Hate speech is neither specifically prohibited under the *Convention on the Prevention and Punishment of the Crime of Genocide* (1948) … nor under the *Rome Statute of the International Criminal Court*.

He continues to state that:

> *A distinct and particularly egregious form of hate speech, namely, direct and public incitement to commit genocide, represents a recognised exception. It is firmly established as an international crime under the Genocide Convention and the Rome Statute as well as under customary international law.*

So whereas there are limitations in the South African constitution on freedom of speech, international law is silent on hate speech, apart from where hate speech is inciting genocide. This particular stance of international law is meaningless unless there is domestic action against genocidal hate speech. No genocidal hate speech has thus far been criminally prosecuted in South Africa.

South African Criminal and Statutory Law

Apart from the constitutional prohibitions on freedom of speech, South African Criminal Law and statutory law declares incitement to commit crime as illegal. For example, the Riotous Assemblies Act 17 of 1956 states that:

> *Any person who ... incites, instigates, commands or procures any other person to commit any offence, whether at common law or against a statute or statutory regulation, shall be guilty of an offence and liable on conviction to the punishment to which a person convicted of actually committing that offence would be liable.*

If a person addresses, for example, a crowd and uses words which incite people to illegal activities or such communication is conveyed via the media or other means, it is reasonable to conclude that the inciter should have foreseen the possibility that the incitement could result in serious actions by those who heard or read it. This is clearly stipulated in the quotation below:

> *In order to obtain a conviction of incitement, the prosecution need not prove that, as a result of the inciting words, the incitee indeed committed the crime. No causal relationship between the accused's words and any subsequent action by the incitee is required (State v Nkosiyana 1966 (4) SA 655 (A)* (Herman & Snyman 2012).

Data collected from the media should be interpreted within the framework of international law and domestic law. From a legal perspective, some of the media postings reflected in the data presented in this article, would in all likelihood suffice for a conviction based on the above legal criteria. It is also important to examine intent (threats and instructions) as well as wrongfulness in relation to hate speech, as a test for contravention of legislation and precedent.

Intent (Threats and Instructions) and Wrongfulness in Relation to Hate Speech

According to Reddi *in* Mbanjwa (2011:1) there are three different types of intention in South African law, 'direct intention, indirect intention and *dolus*

eventualis, also known as the legal intention'. Proving all three forms of intention can lead to conviction, and according to him in *dolus eventualis*, 'a person subjectively foresees the possibility that in achieving his main aim, the unlawful act may result and he is reckless as to whether or not the unlawful act may be committed'. This becomes significant in examining hate speech, since two major empirical trends were evident in the data, which amounted to instruction (i.e. from politicians) and threats (from the ordinary person) on social media.

The legal conceptualisation of threats is descried as follows: 'A threat is any communication indicating an intention to do harm. It can be communicated directly or indirectly either by words (whether written or spoken) or by conduct, or a combination of both' (Criminal Law Consolidation Acts [SA] c19 [3]). Intimidation has also been described as:

> *A statutory offence is defined in the Intimidation Act No. 72 of 1982. In most cases intimidation occurs when a person uses the threats of violence, or takes any steps in relation to those threats of violence, in order to induce a victim to do something he is legally not entitled to do or to induce the victim to abstain from doing something which he is legally entitled to do* (Meiring 2011).

In the same way we need to look at instruction (to do harm, wrongfulness) in relation to social media. This is interpreted against the definition of incitement (Basse 2017):

> *Another legal instrument is the Promotion of Equality and Prevention of Unfair Discrimination Act of 2000, often referred to as the Equality Act. The definition of hate speech is much broader here, since it includes expressions that are 'hurtful', 'harmful' or that will 'incite harm' or 'promote or propagate hatred'. This is evidence of the problem of defining hate speech.*

Theoretical Orientation from Communication Studies

The theoretical orientation of this study from a Communication Studies perspective assumes that those who foster hate speech online has a certain image to maintain, tries to manifest a certain identity in society online (which may

differ even substantially from the identity of the person in real life) and are posturing, performing and playing to political, religious or some other overtures. Taking these considerations into account, this study therefore conceptualises social media and identity formation, impression management online, fragmented internet identities against the background of Goffman's dramaturgical theory and Identity management theory. This needs to be viewed in relation to the Spiral of Silence theory, which needs to be seen as a framework for drawing broader conclusions in relation to social media and hate speech.

Social Media Identity Formation and Impression Management

As far back as 1959 Goffman concluded that self-presentation is becoming popular and easy due to the use of the new technological developments. Gone are the days where users used to find it difficult to present themselves. The distance between performer and the audience that physical detachment provides makes it easy to conceal aspects of the offline self and embellish the online (Goffman, 1959: 7-117). Halle (1996), state that most of the users prefer the use of online communication because of impression management; they get the chance to decide on what their audiences get to see and know about them. Not all information is revealed about them. In the presentation of the self, online users may select certain behaviours and characteristics, as well as using pseudonyms and false or even hacked identities (identity theft) in order to create a desired impression. Baptista (2003) found that new identities are not created online only, they can also be found in real or in everyday face-to-face interaction. In the case of hate speech, the phenomenon is driven by political parties, political beliefs and political posturing. The examples given in the media by politicians create the impression that it is acceptable to behave in this way, and in some cases politicians get away with it. Hence social media users emulate these behaviours, such as hate speech, which manifest online.

Fragmented Internet Identities

With different social media platforms available in the cyber domain, in the 21st century, the number of the users has also increased rapidly from 12% in 2005 to 90% in 2015 according to a research study conducted by the Pew Research Centre (2015). The research study on 'Fragmented Selves: Temporarity and Borderline Personality Disorder' by Thomas (2000), found that the people who

have fragmented identities are suffering from personality disorders, since they constantly change their identities when interacting with different people. Goffman (1959) states that the people who have different identities mask themselves because they interact with different people on different media platforms, so they must meet the standards and requirements of being in those groups, hence they have different identities. It is thus very important to ensure that you present yourself well to people; this can only be achieved through conducting research beforehand. In an article entitled 'the study on the presentation of self in the online world: Goffman and the study of online identities', Bullingham *et al.* (2013) found that users have different identities, the reason being that they must maintain different blogs and different social media platforms so they must create different personas to suit each blog.

Goffman's Dramaturgy and Identity Management Theory

Goffman's dramaturgical approach is used to explain how people present themselves. Goffman (1959) states that life is a stage where individuals engage in performances, with individuals being performers. There are also observers who observe what other performers do and adopt the way they do things. When an individual plays a part he implicitly requests his observers to take seriously the impression that is fostered before them (Goffman 1959: 17). In the cyber, online environment, users are constantly trying to present what they believe the observers should see. Thus, an individual need to adjust their public images with the expectations of their target audience. It is very important to have information about an individual or group before interacting with each other because the information helps to define the situation, enabling others to know in advance what could be expected from them and what they may expect of the user. The *stage* in this study would be an online environment where all the users interact with each other. According to Goffman (1959), the behaviour and actions of the users is actually determined by the target audience. People are performing roles on multiple stages simultaneously, with a globally distributed range of actual and potential audiences (Markham 2012:2). Markham used the dramaturgy approach and Gergens (1991) notion of the saturated self in his study. This study proposes that people have fragmented identities when interacting with other (fragmented) identities.

Identity management theory (IMT) is an intercultural communication theory that was developed by Cupac, Tadasu and William in the 1990s, which

explains the ways in which people handle various situations they find themselves in. Identity management theory was developed to support Goffman's interactional ritual face-to-face behaviour; but took a different direction and looked at the speakers coming from different cultures in real time communication. This study will not look at face-to-face interaction, but examines an online environment while looking at communicators coming from the same background and users from different cultural backgrounds. According to Cupac *et al.* (1990), IMT focuses on how people from different backgrounds, whether that is religion, ethnic, or social backgrounds manage their identity when communicating with other people from different backgrounds. In online environments it is mostly people from different backgrounds who engage in hate speech, thus making this theory relevant. IMT suggests that the message initiators are the ones who decide upon communication behaviour and how others perceive them. Identity management theory involves several key concepts, including competence, identity, cultural and relational identities, face and facework. Cupac, *et al.* (1990), went further to say that effective identity management is very important when interacting with different people and it can only be achieved using identity management theory.

The Spiral of Silence Theory

The Spiral of Silence Theory was first proposed by Noelle-Neumann in 1974, and is described by Liu and Fahmy (2011) in Malaspina (2014) as follows:

> In the 1970s, Elizabeth Noelle-Neumann developed a theory that suggested that the expression and formation of public opinion results from people's perception of the climate of opinion. Individuals use a 'quasi-statistical sense' to determine whether their opinions are popular or unpopular. If they perceive that they share their opinions with the majority, they may be willing to speak out. Alternatively, if they perceive their opinions to be those of the minority, they will keep silent or conform to the majority view.

In a study conducted in 2014, Malaspina concluded that in the social media environment the spiral of silence manifests in the following way:

> The combination of perceived empowerment, strong negativity and

> aggressiveness that has emerged from the findings reflects the bottom-up approach made possible by social media, and the increasing readiness of online users to speak out on controversial topics, as opposed to the top-down structures promoted by institutions, in which individuals are passively subject to the influence of mass media.

It can be seen from the above that, whilst the theory is premised on the idea that people will remain silent, within a social media environment, they will be more likely to speak out and let their opinions be heard. This amounts to a decrease in the perceived size of the spiral, and knowing that a decrease in the size of the spiral is likely to cause dissention, social media thus becomes an agitator, a medium that is *more* likely to cause differences between people to manifest.

Methodology

The research is a desk top study based in the qualitative paradigm having the aim to thematically analyse (c.f. Nowell *et al.* 2017) online data of perceived hate speech and to analyse the text to determine whether such online posts are in contravention of the limitations of sub-section 16(2) of the Constitution. Data were collected from the internet and newspaper articles, quoting posts expressing racial, religious and violent overtures. The analysis in particular searched for texts that expressed content of:

a. Propaganda for war;
b. Incitement of imminent violence; or
c. Advocacy of hatred that is based on race, ethnicity gender or religion and that constitutes incitement to cause harm.

The *verbatim* posts, are quoted without the facebook, twitter or other online identity references, and ensure anonymity (although the data are in the public domain). The purpose is not to expose individuals but rather to analyse text and context within the framework of the selected theories and legal prohibitions. The data were gleaned from 33 online posts and within the three prohibitions.

The steps followed in this study constituted of first examining a sample of ‘viral cases’ of hate speech which occurred in social media. Thus, in terms

of being included in the sample, the identified case of hate speech had to have manifested in some form of social media. The material was sampled in terms of two cohorts.

In the study it was the intent to examine, first, specifically cases that concerned 'the ordinary person' (1st cohort). The notion of the ordinary person concept here was that many of the viral cases that occurred, was *not* uttered by politicians or people that have any fame in the political or *any other* arena whatsoever. It was 'ordinary people' (i.e. policemen, factory workers, plant foreman, etc.) who have *no elevated identity* that would normally be associated with hate speech in the world outside the cyber domain. The occurrence of hate speech could thus be attributed to occurring in social media, wherein, previously referred to as *dramaturgy* and *fragmented identities* in terms of the literature review, it was clearly evident. Criminologically, the availability of the cyber domain, seems to lure ordinary people to post opinions on social media that they probably would hesitate to do before an open audience.

The second cohort specifically examined cases of politicians and influencers of public perceptions, cases that were not necessarily viral in social media, but manifested itself in social media as a result of a hate speech being advocated in some *other* platform (i.e. a speech made by a politician at a political rally) which then was further redistributed through social media. What was important about this cohort was that it did not represent *the ordinary man*, but politicians and other famous figures.

Data from the two separate cohorts were examined separately and analysed by searching for emerging themes. The themes are discussed individually in the next section, the presentation of findings. In relation to the conclusion, the themes are associated with the theoretical orientation of fragmented identities and Goffman's dramaturgy, which concludes in a theoretical model that can be used for the understanding, interpretation and analysis of hate speech.

Presentation and Discussion of Data

Data are presented under the prohibitions of section 16 (a; b; c). Significant quotes from the data are presented in as far as they contravene one or more of the prohibitions and also to indicate by which cohort (1st cohort - the ordinary person; the 2nd cohort - politicians) the cyber posts have been made. There are posts that contain more than a contravention and therefore themes are also

presented under each prohibition. The discussion addresses both the criminological contraventions as well as communication – related aspects detectable from the data. The following themes emerged from the data:

Propaganda for War

> *If your opinion paid any of my bills white boy I'd give a s**t. You and the rest of your oppressing kind need to thank God every night before going to bed that we haven't unleashed hell on you. But fear not a civil war is on the horizon, black people as a collective have realised that we really have nothing to lose whether the economy is good or bad because you monkeys still control through generations of theft … (ordinary person).*
>
> *'Land reform needs an act as forceful as war, and that it was an illusion that land would be given back to black South Africans peacefully'… (politician).*

Only two of all of the posts examined made reference to war. It brings the question to the fore whether statements that the first limitation deals with imminent violence, the authors opted to classify calls for violence against a particular group into that category. The above posts amount to an unambiguous contravention of subsection 16(1). The first post is in contravention of sub-sections (a) and (c):

a. Propaganda for war - *a civil war is on the horizon (ordinary person)* and *Land reform needs an act as forceful as war (politician)*
c. Advocacy of hatred that is based on race, ethnicity gender or religion and that constitutes incitement to cause harm - *If your opinion paid any of my bills white boy I'd give a shit… you monkeys still control through generations of theft… (ordinary person).*

These statements are unambiguous in their intention, is propagating for a factional war as accepted by the authors from Thefreedictionary.com (2018:1). The latter, inter alia defines war as, '…conflict between nations or factions within a nation ...'. Secondly the data correlates with Baptista's (2003)

findings, namely that new identities are not created online only, and that they can also be found in real or in everyday face-to-face interaction. In the case of hate speech, the phenomenon is driven by political parties, political beliefs and political posturing. The rhetoric that whites have stolen the land is also clearly construed. Public statements are frequently made by politicians to this effect.

Incitement of Imminent Violence

They like stupid animals. We should tie them to a rope. Too many Africans flocking to Hout Bay. Draw up a petition. Soon there will be nothing left of Hout Bay … (ordinary person).

I want to cleans (sic.) this country of all white people. We must act as Hitler did to the Jews … (ordinary person).

So, these people, when you want to hit them hard – go after a white man. They feel a terrible pain, because you have touched a white man … (politician).

Most of the social media posts fit into this category and emerged as an overarching theme that were related to all the other themes. Whether the theme post related to infanticide/children, religion and gender, incitement for war, the element of incitement seemed to be an essential element of hate speech on social media. The posts also contained direct incitement to violence against all race groups. Violence are incited against all races, black, white, 'boers', farmers, Indians and Nigerians. Some of the posts contains threats to women and infants. Some posts are cited to demonstrate the incitement to violence but also the direction of that violence is intended.

From a criminological point of view, the data contravenes all of the aspects relating to the limitation to freedom of expression. What is clearly and easily detectable from the data is the fact that incitement to genocide are present in many posts, as can be seen in the posts listed above. As per the theoretical discussion, incitement to commit a crime is illegal, and intent in all of the posts are direct and can lead to direct consequence. The data also presented a very clear distinction between the 1st (ordinary person) and 2nd

(politicians) cohort, in the sense that, in the majority of cases, the ordinary person seems to be mostly posing a threat (i.e *we should…*) whilst politicians seem to be giving instruction (*go after the white man*).

From a communication theory perspective we can clearly see how the dramaturgy plays out, it seems as if the more dramatic the post by the ordinary person the more satisfied they seem to be. The demonstration of this manifested in a number of posts:

> *… screenshot this comment and send it to all your relatives and make sure to bring it up at the next Whites-Only gathering … (ordinary person).*

This tendency was however not as clearly detectable in the politicians cohort. In the politicians cohort the distinct impression is gained that social media becomes more of an impression management tool, and that the comments posted on social media emulates what was iterated on other communication platforms, such as parliament, speeches at rallies, etc.

The impression is also created that the manifestation of fragmented identities clearly takes place. It is the ordinary person (the estate agent, the steel worker and the teacher) who states these comments which then become viral. One gets the impression that these ordinary people did not expect their comments to go viral, as if they were 'caught out' by the impact and veracity of the medium. The medium provides an outlet to ordinary happenings in everyday life:

> *F****ing k*** taxi. And once again I vote for the death penalty. These savages don't deserve to live. But more importantly Daniel is alive and I am alive. They can rot in hell … (ordinary person).*

In the above instance it can be seen that the person was clearly in some sort of a motor vehicle incident during the day and expressed her unhappiness on social media. The problem with the medium is that it allows for the post to become viral, to allow people to contravene legislation in the public sphere whilst in some instances not realising the impact of their communication as

well as the extent to which existing legislation is contravened. Can the ordinary person be expected to have an understanding of the ability of social media to 'go viral' and have an extensive understanding of what constitutes criminality in relation to hate speech and the incitement of violence?

Infanticide / Children

> *… Kill all white men, white women, white babies, white blind and cripple crakkas, white… (ordinary person).*
>
> *We have no choice but to kill the white babies, simply because they are goin to grow and oppress our babies, so we kill the white babies… (ordinary person).*
>
> *Muslim people should just bomb themselves mother… why kill innocent children? I can understand you are circumcised and can't get children but… come try me I haven't used bullets in a long time… (ordinary person).*

The references to children and infants were a common manifestation in the data, and appeared in approximately a third of all of the posts examined. The direction of the violence were clearly detectable, amounted to threats since all of the posts came from the ordinary person and none from the politicians.
From a criminological point of view it can be seen that babies and children are mentioned in the post since it adds to the veracity of the post. In South African law violence against children have been well documented as being one the most hideous acts that any person can be found guilty of. The impression is gained that it creates an ultimate form of intimidation, and the incitement thereof is clearly not only illegal, but also extremely distasteful to those in receipt of whom the communication is directed at. The mere fact that none of the politicians (2nd cohort) made themselves guilty of this crime is indicative of the management of communication aspect … from a legal point of view they are clearly well informed about what can and cannot be posted, and are better equipped than the ordinary person.

From a communication studies perspective it can be seen that the incitement of such hideous crimes becomes the ultimate dramaturgy in relation

to Goffman's theory. The higher the level of dramaturgy the higher the level of intimidation, and the higher the level of fear that is potentially evoked, in the victims of the hate speech. The impression is gained that this is the ultimate purpose of the communicative act, instilling fear.

The researchers of this paper are not of the opinion that these posts, especially related to infanticide and children, are posted by people who would actually portray these acts, with the possible exceptions of truly deranged individuals. The researchers are of the opinion that these are ultimate manifestations of fragmented identities that is described in the theoretical section of this paper. From a political point of view thus, the person making these posts would make these statements online to create fear, to create an image of ultimate disrespect for human life, whilst the real life identities are far removed from the identities that are portrayed online. In terms of the spiral of silence theory people have over time learnt that their comments can create an effect, and as to whether their acts are acted upon (or not) become irrelevant in relation to the effect that they have on society.

Religion

White S Africans your time is up. In the near future you will pray and wish you were black. Your time is up. Pack and go or suffer the consequences. God is not on your side. The bible you brought is working against you. Blacks are awake. You better note that ... (ordinary person).

Muslim people should just bomb themselves mother… (ordinary person).

You are classified in the Bible as an animal, you are not homosapien (sic.)… (ordinary person in radio interview following racist tweet for refusing black people to use his guest house).

Posts related to religion amounted to approximately a quarter of all posts examined. Religion, in the same manner as infanticide and harm to children, can be seen as one of those core beliefs that people hold dearly, and in the process of instilling fear, becomes a target for criminals using the digital communicative agent upon their victims.

From a criminological point of view it can be seen that freedom of religion is seen as one of those inalienable human rights that is viewed as so sacred that it is considered a basic human right. The crime of these posts are related to section 16(2) of the Constitution which states that 'Advocacy of hatred that is based on race, ethnicity gender or religion and that constitutes incitement to cause harm'.

As with the previous theme, from a communication theory point of view it can be seen again that Goffman's dramaturgy is taking place. Of interest here is the manner in which a guesthouse owner, who were investigated by the Human Rights commission for stating online that he refuses to allow black people to stay in his guesthouse, took to traditional media and repeated over radio the post mentioned above. The fact that this occurred is different from most other cases examined in this cohort where viral comments were apologised for once they became viral, such as:

> *I'm sorry to say I was amongst the revellers and all I saw were black on black skins what a shame. I do know some wonderful thoughtful black people.*

Of deeper concern here is the *extent* to which religion is insulted (c.f. above). Whilst a lot can be said about online dramaturgy, fragmented identities and the spiral of silence theory and its applicability to this study, it should be questioned whether some individuals are prepared to incite to the extent that they (clearly) have no care for the effect that it is having upon the victims of the communicative act.

Findings and Conclusion

From a criminological perspective the study found that the examined posts from South African users of cyber space were in contravention of International Law as contained in the prohibitions on hate speech as contained in *Convention on the Prevention and Punishment of the Crime of Genocide (1948)* and sub-section 16(2) of the Constitution. Furthermore, statutory and common laws were violated such as the Riotous Assemblies Act 17 of 1956. The study found that the posts examined, in the understanding of the legal context have used the cyber domain to contravene international law, i.e. *'Kill all white men, white*

women, white babies, white blind and cripple crakkas, white …' (ordinary person).

Incitement of violence emerged as a common theme. This is both unconstitutional and against common law and statutory law. The Riotous Assemblies Act 17 of 1956 states that *'Any person who ... incites, instigates, commands or procures any other person to commit any offence, whether at common law or against a statute or statutory regulation, shall be guilty of an offence …'.*

Posts that propagate hate and violence were based on race, gender, religion and even against infants, propagating infanticide, i.e. *We have no choice but to kill the white babies, simply because they are going [sic] to grow and oppress our babies, so we kill the white babies. Violence are incited against all races, black, white, 'boers', farmers, Indians and foreign nationals like Nigerians and Zimbabweans.*

Justification were found in past injuries, especially the racial policies of apartheid, economic inequality, land ownership and racial superiority. The equating of people to 'monkeys' and stealing the land were typical examples. Religion was found to be both a motivator for instigating violence and a justification thereof.

In terms of intention it was found that all forms of intention were present in many of the posts (direct, indirect and *dolus eventualis*), which are sufficient to lead to conviction. Two major empirical trends were evident in the data, which amounted to instruction (i.e. from politicians) and threats (from the ordinary person) on social media. These threats, defined as 'any communication indicating an intention to do harm can be communicated directly or indirectly either by words (whether written or spoken) or by conduct, or a combination of both' (Criminal Law Consolidation Act (SA) s19 (3)), clearly occurred in most of the social media posts. Intimidation, as defined in the Intimidation Act No. 72 of 1982, which is a type of threat, also occurred in most posts examined from the data. In terms of the Promotion of Equality and Prevention of Unfair Discrimination Act of 2000, wherein a broader definition of hate speech is defined, hate speech was also evident and posts were clearly in contravention of the prohibitions of the Act.

From a communication theory perspective, the theories mentioned in the literature review manifested in a unique and very specific manner. If Goffman's dramaturgical approach is examined it can be seen that the South African posts were *extremely* dramatic in their approach to social media, to the

extent that contravening the law in relation to hate speech posed no problem. It went further, to the extent that core belief systems were attacked, such as religion, and that infanticide and harm of children posed no problem for these individuals. Incitement of violence and propaganda for war were the most extreme forms of dramaturgy encountered.

The question needs to be asked whether these dramaturgical presentations on social media are reflective of true intent and whether the threats and intimidation would all be acted out by the individuals who posted them. By examining posts made by the ordinary person whilst considering the criminological orientations of intent, it is the opinion of the authors that, in many instances, especially where the outcomes of court cases are known, these individuals were surprised by the fact that their posts became viral, let alone would have led to a conviction at the time of posting them. Whilst the dramaturgy is clearly evident, an online and fragmented identity thus emerges. The identity is fragmented because it does not reflect the true convictions of the person, it is an online identity which is very different from the known identity of that person in everyday life. The ordinary person is not knowledgeable about the law and its ramifications, and would not hesitate to reflect their anger about a specific situation on social media, which culminates as threats. Since the majority of threats cannot be acted upon due to geographical displacement between people posting on social media, these threats are viewed by both the person making the post and the reader as empty, and difficult to act upon.

In relation to identity management theory it can clearly be seen from the extent of the contraventions of the law in relation to hate speech that cultural and relational identities were reflective of what was happening on the political scene in South Africa. The fact that ordinary people's posts mostly amounted to threats and that politicians' posts amounted to instruction are significant here. Some politicians in South Africa publicly have no reservations to publicly expressing hate speech (which leads to investigations and convictions by bodies such as the Human Rights Commission) and these sentiments are emulated by social media users and manifest online. Identity management theory postulates that people manage their identity online in a specific manner when they encounter people from different cultures and backgrounds. In the context of this study this definitely occurred, and showed how identities were managed in a particularly negative context, and that people react in a specific manner when encountering people from different backgrounds and cultural identities in relation to hate speech.

In terms of the spiral of silence theory, the above is exacerbated by Malaspina's (2014) conclusion that in the social media environment the spiral of silence manifests by the *'combination of perceived empowerment, strong negativity and aggressiveness that has emerged from the findings reflects the bottom-up approach made possible by social media'*. It can thus be stated that from this study that it can be concluded that social media has a negative effect on people's tendency to remain silent and follow popular opinion, and that it exacerbates the manifestation of all forms of communication, inclusive of hate speech.

In conclusion, this article demonstrated that in relation to hate speech, and the manifestation thereof on social media, South Africa has a very worrying tendency to contravene all legislation related to hate speech, nationally and internationally. This manifested to the extent that propaganda for war and incitement of violence appears frequently in the media. This gets done in laudatory terms, is extremely dramaturgical, and quite often emulates the examples set by politicians in traditional media. The specifics of the medium of social media exacerbates the situation, due to the fact that social media leads to fragmented identities and poor management of identity when encountering people from different race groups. The medium does not support people to remain silent and follow popular opinion, but rather express their views which quite often is in contravention of South African legislation. In understanding the cyber mind in relation to hate speech, not only do criminologists, legislators and the legal fraternity need to bear in mind that there is a high frequency of contravention, but that the communicative aspects of the cyber medium lends itself to the fostering of hate speech and the manifestation thereof. Higher levels of sensitivity are thus required, strict application of the law and communicological concepts such as fragmented identities and medium specifics need to be constantly borne in mind.

References

Agboola, A. 2018. South Africa Votes to Take Back 'Stolen' Land from White Owners. Available at: http://www.blackenterprise.com/south-africa-votes/ (Accessed on 19 April 2018.)

Baptista L. 2003. Framing and Cognition. In Trevin, A (ed.): *Goffman's Legacy*. Lanham, MD: Rowman & Littlefield.

Basse, K. 2017. Hate Speech on Social Media: A Comparison between South Africa and Germany. Available at: https://kristinbsse.atavist.com/hate-speech-on-social-media (Accessed on 01 April 2018.)

Boggenpoel, Z.T. 2013. The Prosecution of Incitement to Genocide. *South Africa [2013] PER 74 Potchefstroom Electronic Law Journal // Potchefstroomse Elektroniese Regsblad*. Available at: http://www.saflii.org/za/journals/PER/2013/74.html (Accessed on 01 April 2018.)

Bullingham, L. & A.L. Vasconcelos 2013. The Presentation of Self in the Online World: Goffman and the Study of Online Identities. *Journal of Information Science* 39,1: 101 - 112. https://doi.org/10.1177/0165551512470051

Businesstech. 2016. White people don't own the land – so take it: Malema. Available at: https://businesstech.co.za/news/government/128169/white-people-dont-own-the-land-so-take-it-malema/ (Accessed on 19 April 2018.)

BusinessTech 2017. White people didn't steal the land – they bought it. Lekota. Available at: https://businesstech.co.za/news/general/169737/white-people-didnt-steal-the-land-they-bought-it-lekota/ (Accessed on 19 April 2019.)

Cupach, W.R. & T.T lmahori 1990. Intercultural Communication Competence: Culture-general, Culture-specific, and Culture-synergistic. Paper presented at the annual meeting of the Speech Communication Association, San Francisco.

Cupach, W.R. & T.T. Imahori 1993. Identity Management Theory: Communication Competence in Intercultural Episodes and Relationships. Wiseman, R.L. & J. Koester (eds.): *Intercultural Communication Competence*. Newbury Park, CA: Sage.

Gergen, K.J. 1991. *The Saturated Self: Dilemmas of Identity in Contemporary Life*. New York, NY, US: Basic Books.

Goffman, E. 1959. *The Presentation of Self in Everyday Life*. Garden City, NY: Doubleday Books.

Halle, D. 1996. *Inside Culture: Art and Class in the American Home*. Chicago, IL: University of Chicago Press.

Herman, D. & C.R. Snyman 2012. Why Malema can be charged with incitement – Solidarity. Available at: http://www.politicsweb.co.za/news-

and-analysis/why-malema-can-be-charged-with-incitement--solidar (Accessed on 01 April 2018.)

Malaspina, C. 2014. *The Spiral of Silence and Social Media: Analysing Noelle-Neumann's Phenomenon Application on the Web during the Italian Political Elections of 2013.* MSc in Media, Communication and Development, Department of Media and Communications, London School of Economics and Political Science: London.

Markham, A. 2012. Dramaturgical Approach: What's different about digital experience? Available at: https://annettemarkham.com/2012/02/dramaturgy-and-digital-experience/ (Accessed 1/4/2018.)

Mbanjwa, B. 2011. Cleaning up Confusion over Intention. *Daily News* 20 December. Available at: https://www.iol.co.za/dailynews/news/clearing-up-confusion-over-intention-1201451 (Accessed on 02 April 2018.)

Meiring, H. 2011. Dealing with Threats of Violence. *GOLEGAL Industry News and Insight.* Available at: https://www.golegal.co.za/dealing-threats-violence/ (Accessed 27 March 2018.)

Noelle-Neumann, E. 1974. The Spiral of Silence: A Theory of Public Opinion. *Journal of Communication* 24,2: 43 - 51. https://doi.org/10.1111/j.1460-2466.1974.tb00367.x

Nowell, L.S., J.M. Norris, D.E. White & N.J. Moules 2017. Thematic Analysis: Striving to Meet the Trustworthiness Criteria. *International Journal of Qualitative Methods* 16,1:1 - 13. https://doi.org/10.1177/1609406917733847

PewResearch Centre 2015. Teens, Social Media & Technology Overview: Smartphones Facilitate the Shifts in Communication Landscape for Teens. Available at: www.Pewresearch.org (Accessed on 12 March 2018.)

Statista.com 2018. Number of Monthly Active Facebook Users Worldwide as of 4th Quarter 2017. In Millions. Available at: https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/ (Accessed on 02 March 2018.)

Statista.com 2018. Number of Monthly Active Twitter Users Worldwide from 1st Quarter 2010 to 4th Quarter 2017. In Millions. Available at: https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/ (Accessed 02 April 2018.)

The English Dictionary 2018. Available at:

https://www.collinsdictionary.com/dictionary/english/war (Accessed on 25 March 2018.)

Thefreedictionary.com 2018. Available at: https://www.thefreedictionary.com/war (Accessed on 25 March 2018.)

Thomas, A. 2000. Textual Constructions of Children's Online Identities. *CyberPsychology and Behaviour* 3,4: 665 - 672. https://doi.org/10.1089/109493100420250

Ian P. Saunderson
Communication Studies
University of Limpopo
ian.saunderson@ul.ac.za

Cornelis Roelofse
Faculty of Law
University of Zululand
Mtunzini

Christopher Gumbi
Department of Correctional Services
Tshwane