



SEEK WISDOM, ELEVATE YOUR INTELLECT AND SERVE HUMANITY !



Typology and Mechanisms of Ethnic-based Online Hate Speech in Ethiopia: A Grounded Theory Approach

By

Tadesse Megersa Shiferaw

Addis Ababa University

Addis Ababa, Ethiopia

June, 2024

**Typology and Mechanisms of Ethnic-based Online Hate Speech in
Ethiopia: A Grounded Theory Approach**

A Dissertation

Submitted to the School of Psychology of Addis Ababa University in Partial
Fulfillment of the Requirements for the Degree of Doctor of Philosophy in
Applied Social Psychology

By

Tadesse Megersa Shiferaw

Advisor

Abebaw Minaye (PhD, Associate Professor)

School of Psychology

College of Education and Behavioral Studies

Addis Ababa University

June, 2024

Addis Ababa, Ethiopia

Declaration

I, Tadesse Megersa, hereby declare this dissertation work with the title ‘Typology and Mechanisms of Ethnic-based Online Hate Speech in Ethiopia: A Grounded Theory Approach’ which is submitted by me to the School of Psychology of Addis Ababa University for the partial fulfillment of Doctor of Philosophy in Applied Social Psychology is my original work, and has not been previously submitted by me or anybody else at another University. The materials obtained from other sources have been duly acknowledged in the document.

Candidate’s Name: Tadesse Megersa ***Sign:*** _____

School of Psychology

College of Education and Behavioral Studies

Addis Ababa University

I hereby certify that I have read the revised version of this dissertation entitled ‘‘Typology and Mechanisms of Ethnic-based Online Hate Speech: A Grounded Theory Approach’ which is prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Abebaw Minaye (PhD, Associate Professor, Advisor): _____ ,

Signature

Date

Abstract

Using language that devalues others is known to be the royal road to enemy-making, identity based conflict, and genocide. Parallel to the explosive growth of social media users in Ethiopia, ethnic-based hate speech has increased to the point where it is ripping Ethiopian society apart and threatening the values of democracy, human dignity, and peaceful coexistence. The present grounded theory-inspired study aims to explore the types of ethnic-based hate speech, how it occurs, and how social media users react to hate speech posts on social media. In order to address these objectives, relevant social media contents are collected from a few purposefully selected social media accounts or pages, from Facebook, YouTube news Channels, and Twitter platforms. Using a broader working definition and a checklist, hate speech data was manually collected, while comments given to a few of the hate speech posts were automatically extracted using a web link called exportcomment.com. In addition to social media content, six interviews were also conducted to support the findings obtained from the analysis of social media content. The collected data are stored in dataset, one by one until data saturation is reached. After immersion into the dataset, the analysis is guided by inductive coding as well as relevant grounded theory principles such as constant comparative analysis. Through inductive coding, hence, codes, concepts, and categories are generated and presented with thick descriptions. The finding shows four types of hate speech (dehumanization, enemification, devaluation, and attacking and repelling), three core categories of mechanisms (earned reputation, trolling and poaching, and user-fallibility), and five roles of users in reacting to hate speech posts (trolling, pacemaking, peacemaking, guarding, and informing). In addition, the findings show four interrelated contexts in which the hate speeches were framed: competition and disagreement over history and resources, the conflict in the Northern part of Ethiopia, lack of diversity tolerance, and failure of the government to uphold law and order. The study discusses the findings and provides recommendations deemed necessary to counteract hate speech.

Keywords: Hate speech, dehumanization, enemification, devaluation, trolling, pacemaking, poaching, earned reputation

Acknowledgements

First and for most, I wish to express my deepest gratitude to my advisor, Abebaw Minaye (PhD., Associate Professor), for his valuable guidance, remarkable patience, and continuous support during my doctoral journey. I take a bow before you for holding my hands through the darkest and the brightest times of this research and making my efforts meaningful. I am also indebted to Professor Zenebe Beyene and Dawit Mekonnen (PhD) the external examiner and the internal examiner for this dissertation respectively. Your constructive comments and encouraging feedback have helped me and the document a lot. I would also like to thank all the examiners who provided critical comments on the proposal and the pilot work of this study (please accept my apology for not mentioning your names).

Second, my heartfelt thanks also go to the Institute of Peace and Security Studies of Addis Ababa University, as well as the French Centre for Ethiopian Studies for offering a data collection grant for this dissertation. Without such generous offers, this stage would have been unthinkable. Third, I am extremely grateful to my key informants, at Ethiopian Media Authority and Center for Democracy and Right, who generously contributed their time and insights. The experience and perspectives you shared with me has provided significant depth to my research.

The initial findings of this study were presented at various national conferences, where constructive comments were obtained. I am therefore thankful to those universities for the oral presentation I enjoyed and the individuals who provided me with critical comments, thought stimulating questions, and genuine encouragement.

Last, but not least, I would like to extend my loving gratitude to my wife, Mahlet Teshome. Your love, understanding, and encouragement you offered me have been the bedrock of my strength in this work. While juggling between my research and our home, your presence made it all bearable and worthwhile. Our kids Noah and Yoadan too deserve the credit for inspiring and entertaining me at home, where much of this work is done (but I need their apologies for the reaction I showed them when they, sometimes, disturbed me). In conclusion, the successful completion of this dissertation would not have been possible without each one of you. While these words are hardly enough, this dissertation is a testament to your immeasurable contributions, an accomplishment I share equally with you all. Thank you!

Table of Contents

Abstract-----	i
Acknowledgements-----	ii
Table of Contents-----	iii
Lists of Figures-----	viii
Lists of Tables -----	ix
Acronym -----	x
Technical Terms-----	xi
Prologue-----	xii

Chapter One: Introduction

1.1 Background of the Study-----	1
1.2 Statement of the Problem -----	5
1.3 Research Questions -----	11
1.4 Significance of the Study -----	11
1.5 Scope of the Study -----	13
1.6 Definition of Key Terms -----	13

Chapter Two: Literature Review

2.1 Concept of Hate: an intergroup aspect -----	15
2.2 Duplex Theory of Hate-----	17
2.2.1 Triangular Theory of Hate-----	17
2.2.2 Hate as a Story-----	18
2.3 Causes of Intergroup Hatred-----	19
2.3.1 Psychodynamic View-----	20
2.3.2 The Self: Identity and Intergroup Hatred -----	20
2.3.3 Cognition in Hatred-----	23
2.3.4 Sociopolitical and Historical Contexts of Hatred-----	26
2.4 Hate Speech: The Language of Hatred-----	28
2.4.1 Conceptualizations of Hate Speech-----	28
2.4.2 Causes of Hate Speech -----	32
2.4.2.1 Individual Level Factors-----	32
2.4.2.2 Narratives and Politics-----	33
2.4.3 Consequence of Hate Speech-----	34

2.4.4	Actors in Hate Speech and their Techniques-----	39
2.5	Social Media: <i>(Mis)use and Hate Speech</i> -----	40
2.6	Social Media Theories -----	43
2.6.1	Theories from Media and Communication Fields-----	44
2.6.2	Social Psychological Theories of Media Use-----	46
2.6.3	Personality and Social Media Use -----	47
2.7	Typologies of Social Media Users-----	47
2.8	Countering Hate Speech-----	49
2.8.1	Legal Means -----	49
2.8.2	Educational Means-----	51
2.8.3	Working with Media Platforms -----	52

Chapter Three: Methods

3.1	Study Design: A Grounded Theory Study-----	54
3.2	Philosophical Assumption -----	56
3.3	The Research Road Map-----	57
3.4	Sources of Data -----	58
3.4.1	Social Media-----	58
3.4.2	Key Informant Interviewees-----	61
3.5	Data Collection Procedures -----	62
3.5.1	Judging Contents as Hate Speech-----	62
3.5.2	Inclusion Criteria-----	62
3.5.3	Time Frame for Data Collection-----	63
3.5.4	Data Collection Techniques -----	64
3.5.4.1	Manual Searching -----	64
3.5.4.2	Automatic Searching -----	66
3.6	Theoretical Saturation -----	66
3.7	Dataset Formation and Data Clearance-----	67
3.8	Data Analysis-----	67
3.8.1	Memoing-----	68
3.8.2	Familiarization and Context Identification-----	68
3.8.3	Coding-----	68
3.9	Trustworthiness-----	71
3.10	Human Subject Protection-----	73
3.11	Researcher’s Background, Role, and Possible Bias-----	74

Chapter Four: Findings

4.1 What are the Types of Hate Speeches Circulated on Social Media-----	78
4.1.1 Dehumanization -----	78
4.1.1.1 Animalization-----	78
4.1.1.2 Objectification-----	84
4.1.1.3 Demonization-----	86
4.1.2 Enemification-----	88
4.1.2.1 Land Grabbers, Assimilators, and Settlers -----	89
1.2.1 Unreliable Historical Enemies -----	92
4.1.3 Devaluation -----	93
4.1.3.1 Cheap -----	93
4.1.3.2 Backwards-----	94
4.1.3.3 History-less -----	95
4.1.3.4 Aberration and Conscienceless-----	95
4.1.4 Attacking and Repelling-----	96
4.1.4.1 Direct Call for Attack-----	96
4.1.4.2 Warning and Threatening -----	99
4.1.4.3 Cursing and Ill-thinking -----	99
4.1.4.4 Call for Distancing and Repelling -----	99
4.1.4.5 Calls for and Expressions of Indifference-----	100
4.1.5 The Link among the Types of Hate Speech -----	101
4.1.6 Contexts for Hate Speech -----	103
4.2 How Do Social Media Users React to Ethnic-based Hate Speech Posts? (Typology of users with regard to responding to hate speech posts)?-----	105
4.2.1 Trolling -----	105
4.2.2 Pace-Making-----	109
4.2.3 Peace-Making -----	111
4.2.4 Informing -----	113
4.2.5 Guarding -----	115
4.3 How do Ethnic-based Hate Speeches Occur on Social Media? (mechanisms involved in the occurrence hate speech) -----	119

4.3.1	Earned Reputation -----	120
4.3.1.1	Offensive Episodes -----	120
4.3.1.2	Textual Hate Speech -----	123
4.3.1.3	Liabile Subgroup-----	125
4.3.2	Poaching and Trolling-----	128
4.3.2.1	Accusation -----	130
4.3.2.2	Irritation -----	132
4.3.2.3	Deception -----	134
4.3.2.4	Playing Credible -----	135
4.3.3	Users Fallibility -----	136
4.3.3.1	Random Reaction-----	136
4.3.3.2	Selective Reactions-----	137
4.3.3.3	Tolerance to Own Trolls-----	138

Chapter Five: Discussion and Conclusion

5.1	Discussions-----	139
5.2	Conclusions -----	162

Chapter Six: Implications and Recommendations

5.3	Recommendations-----	164
-----	----------------------	-----

References

Appendices

Appendix 1 Interview Guide

Appendix 2 Sample Transcript of a Hate Speech Video

Appendix 3 Sample Comments Analyzed

Appendix 4 Sample Pictorial Posts Analyzed

Appendix 5 Unstructured Sample Memos

Appendix 6 Hate Speech Screening Checklist

List of Figures

Figure 1 Sternberg's Triangle of Hate -----	18
Figure 2 Pyramid of Hate-----	36
Figure 3 Allport's Levels of Prejudice-----	37
Figure 4 Proposed Antecedent-Consequence of Hate -----	38
Figure 5 The Research Roadmap -----	57
Figure 6 Data Sources-----	59
Figure 7 Data Collection Period and Main Tasks -----	63
Figure 8 Coding Template -----	70
Figure 9 Summary of Types of Hate Speeches-----	100
Figure 10 Proposed Links among the Types of Hate Speech -----	103
Figure 11 Proposed Links between the Roles and Occurrence of Hate Speech -----	118
Figure 12 Proposed Conceptual Model of Hate Speech Occurrence -----	119
Figure 13 Subgroup as a Means for Hate Speech -----	128
Figure 14 Poachers Characteristics -----	129

List of Tables

Table 1 Name Calling in Genocide -----	35
Table 2 Hot Spots-----	59
Table 3 Interviewees' Profile-----	61
Table 4 Safety Check to Guard off Bias-----	76
Table 5 Animalization Expressions -----	78
Table 6 Less-than-Animal Expressions-----	84
Table 7 Objectification Expressions -----	84
Table 8 Demonizing Expressions-----	86
Table 9 Enemification Expressions -----	89
Table 10 Devaluation Expressions -----	93
Table 11 Links among Types of Hate Speeches -----	102
Table 12 Sample Reactions of Users to Hate Speech Posts -----	105
Table 13 Sample Venomous Personal Attack-----	107
Table 14 Sample Venomous Group Attack-----	107
Table 15 Sample Stubborn-Obstinate Reactions-----	111
Table 16 Sample Peacemaking Reactions -----	112
Table 17 Negative Replies Each First Role Obtained -----	117
Table 18 Instances of Offensive Episodes-----	121

Acronym

CGT	Constructivist Grounded Theory
EMA	Ethiopian Media Authority
FDRE	Federal Democratic Republic of Ethiopia
GT	Grounded Theory
HSDSPP	Hate Speech and Disinformation Suppression and Prevention Proclamation
IDP	Internally Displaced People
KII	Key Informant Interview
OLA	Oromo Liberation Army
PP	Prosperity Party
QCA	Qualitative Content Analysis
RTLMC	Radio_Télévision_Libre_des_Mille_Collines
SIT	Social Identity Theory
TPLF	Tigray People Liberation Front
UN	United Nations
UNESCO	United Nation Education, Scientific and Cultural Organization

Technical Terms

Conflict in the North	A two-year deadly conflict between TPLF and the Ethiopian Federal Government, from 3 November 2020 to 3 November 2022.
Ereecha	Thanksgiving festival celebrated at the end of the winter by ethnic Oromos.
Fano	Historically, the term referred to irregular soldiers who voluntarily participated in military campaigns. While writing this paper, an armed group names itself Fano though the government names it ‘rebel group’.
Ga*la	Pejorative and derogatory term assigned to ethnic Oromos in Ethiopia.
Junta	A term reserved for combatants and supporters of TPLF during the conflict in the North
Neftegna	Literally it implies one who carries a rifle. In the hate speech saga, it mainly alludes to ethnic Amhara with the implication of ruling over others by force.
Oromuma	A term which many Oromos assume to imply the essence of being Oromo. Yet, some critics use the term as having hidden political agenda in favor of Oromos only
Prosperity Party	The Leading Political Party in the current Ethiopia, led by PM Abiy Ahmed
Qeero	While it traditionally means ‘bachelor’, within the recent political movement in Ethiopia, it implies the struggle of unarmed Oromo youth against what they believed is unjust and oppressive system
Shene	A name the Ethiopian Federal Government use to refer to an armed group operating in Oromia, which names itself as OLA
Woyane	A term referring to TPLF

Prologue

"Words have the power to shape our world, for better or worse" - (Anonymous)

In a diverse nation like Ethiopia, diversity tolerance, and respecting human dignity are of higher importance. Yet, in recent times, the rise of ethnic-based hate speech has emerged as a pressing issue that threatens the social fabric of its people. This dissertation aims to delve into the complexities surrounding this phenomenon, shedding light on the typologies and mechanisms involved in the circulation of ethnic-based online hate speech in Ethiopia. This report, therefore, presents sensitive content which may challenge readers. As we shouldn't add fuel while wanting to fight back against hate speech, I want to say a few words to readers before letting them enter into the crux of the matter!

The report includes hate speech pictures and quotes used to illustrate concepts, not to blame individuals and groups or to cause sense of victimhood on anyone. The evidence presented is intended to serve as testimonies and visual aids to shed light on hate speech in Ethiopia. As a data collector and interpreter, I understand that the hate speech examples presented in this document don't represent any specific ethnic group, and hence no group should be held accountable for those expressions. The quotes and pictures presented in this monograph could be produced by individuals including fake account holders, or by AI, with varying motives. They can also be produced by individuals who have a clear motive to cause social divisions. What is studied here is hate speech contents which are made public, irrespective of who produced them. Nonetheless, as the hate speech content presented are offending, they might challenge you or other readers emotionally. Therefore, I want to draw audiences' attention not to fall prey to hate speech by taking instances presented here as evidence to strengthen animosity between groups. Perhaps the takeaway from this work is that 'as citizens of the same country, we all are victims' and are laying the groundwork for a dangerous world for us and the generation to come. Therefore, it is my fervent hope that readers will approach this paper with an understanding of its purpose as a means to tackle the problem from its source rather than perpetuating further discord. As we move forward from this report, let us remember the power of our words and actions and the impact they have on our society.

Chapter One

Introduction

1.1 Background of the Study

Despite hopes that humans have become more civilized in the twenty-first century, our intergroup perceptions and treatment of others fall far short of expectations, prompting Gordon Allport to say, "We appear to be living in the stone age" (1954, P. xiii). Echoing this and stunned by genocidal acts in Germany, Cambodia, Bosnia and Herzegovina, Rwanda, and elsewhere, Kressel, author of 'Mass Hate: The Global Rise of Genocide and Terror,' was tempted to consider the willingness to kill other human beings including one's neighbor as an inborn characteristic of the human species (2001).

Abundant evidences show many of the atrocities seen throughout history, as well as the various forms of identity-based attacks that we are currently witnessing in the twenty-first century, can be traced back to negative intergroup attitudes and/or hatred (Allport, 1954; Maddrell, 2015; Morrock, 2010; Rieber, 1991). In support of this, the most recent World Economic Forum study on global risks reveals that societal polarization is not only one of the world's top ten current concerns, but also a long-term threat in the coming decade (Zahidi, 2023). Polarization, which is the shift of people into increasingly mutually opposing cliques that have grown apart in their positions, can considerably contribute to a loss of tolerance for the 'other' group, and it is one of the major elements driving the current wave of democratic backsliding unfolding around the world (Barberá, 2020).

Regrettably, millions of lives, billions of dollars, and hopes and trust between groups are lost because of negative intergroup conflicts. According to the Global Peace Index (GPI, 2023), for example, the economic impact of violence globally is around 17.5 trillion dollars, which is equivalent to 12.9% of total world GDP or \$2,200 per person (2023). While trillions of dollars are lost as a result of conflicts, an enormous budget is injected into the economy each year in the hopes of preventing violence and restoring peace. This situation has, therefore, remarkably and compellingly attracted the attention of scholars, governments, and organizations to invest much more in understanding and tackling the problem of violence. For example, goal 16 of the sustainable development goal aims to promote peaceful and inclusive societies (UN, 2015).

Nonetheless, identity-based attacks like genocide and civil war never happen overnight. Rather, they are preceded by negative attitude formation, polarization, and hate speech. In support of this, evidences show that in nearly all genocides, xenophobic attacks, and mass killings based on protected characteristics, the targets are always assigned with disparaging, dehumanizing, and derogatory labels prior to the attack (Article-19¹, 2015; Council of Europe, 2009; Morrock, 2010). Because expression of hatred in actual terms (i.e., physical attack) is legally, morally, and culturally frowned upon, individuals use words to vent out their hatred against their target, commonly known as hate speech. In support of this, renowned authorities in the fields of hatred (Strengberg, 2005), prejudice (Allport, 1954), and genocide (Stanton, 1998) reiterate that hate speech is one way of venting out negative feelings and emotions one holds against its target.

Hate speech, which some call ‘the language of hatred’ (Anstey et al., 2012; Brindle, 2016; Kennedy et al., 2023), is the process and outcome of enemy-making that denigrates individuals, threatens groups, creates a common enemy myth (Post, 2009), and precedes many of the genocides and atrocities that collective human memory can hold (Benesch, 2014; Staub, 1992; Stein, 2016). Hate speech portrays a group in a delegitimized, devalued, and dehumanized manner and paves ways to exclude its target from the community of humankind through the adoption of a specific ideology and the construction of a common enemy myth (Article-19, 2015), which increases the target group’s vulnerability to various forms of violence (Wibke, 2008) and places a serious barrier to full participation in democracy (International Federation for Electoral Systems, IFES, 2018).

Hate speech can focus on any of the social identities or protected characteristics such as race, ethnicity, disability status, sexual orientation, religion or religious beliefs, sex, social class, gender and gender rearrangements socioeconomic status, age, and national origin (Article-19, 2015; Post, 2009). According to HateBase, a web-based application that collects instances of hate speech online worldwide, the majority of cases of hate speech target individuals based on ethnicity and nationality (Winiewski et al., 2017). Furthermore, hate speech is more prevalent online than offline, to the extent of making it difficult to navigate social media without encountering an offensive and dehumanizing image, word, audio, or video (Jakubowicz et al., 2017). Thanks to the fast spread of mobile phone accessibility, cheaper internet connectivity, and ease of creating and using platforms, social media gives hate

¹ a human rights organization that works to defend and promote freedom of expression and freedom of information

mongers the chance to amplify and spread hate speech (Bossetta, 2018; Mercy Corps, 2019; Singer & Brooking, 2018), leaving naïve social media users to believe and share contents without reflecting on what they are or how they can be used (Hargittai, 2010).

Social media exacerbates polarization by giving users the anonymity they need to speak out, allowing opposing parties to disparage and blame one another while fostering discussions among like-minded people (Sunstein, 2009). Indeed, social media rewards identification and connection within a group, but at the expense of intergroup cohesion (Mercy Corps, 2021). Social media is simple to create, inexpensive to distribute, instantly reaches a global audience, and can resurface and (re)gain popularity over time. Algorithms utilized by online firms further contribute to the rise of hate speeches and the intensification of intergroup hatred, by creating media hype (Barberá, 2020; IFES, 2018).

Hate speech, which intrudes into politics, religion, sport, and other sectors (Nadim & Fladmoe, 2016), is a global concern that forced the United Nations to adopt a strategy and plan of action on hate speech in 2019. As of recent years, the flame of hate speech has been noticed wreaking havoc in society and threatening security issues and peaceful coexistence in Ethiopia (FDRE, 2020; Muluken et al., 2021). While scrolling my own Facebook page on one of the data collection days, for example, I came across the following social media comments and posts just within 20 minutes: ‘kkk, the donkey race’, ‘We don't need you because u r banda (traitors) and beggars’; and ‘we have already told you, the herd cannot think as they have no mind’. The situation of hate speech in Ethiopia has even pushed the government to prepare and approve the ‘Hate Speech and Disinformation Suppression and Prevention Proclamation (HSDSPPE)’ in 2020. In its rationale part, this document states, ‘Cognizant of the threat hate speech and disinformation pose to social harmony, political stability, national unity, human dignity, diversity, and equality...’ (FDRE, 2020, P. 1). This legal document clearly shows the government is feeling the pain hate speech is causing in Ethiopia.

The rise of hate speech in Ethiopia is also parallel with the rise of social media and its misuse by minors (Kinfu, 2017). This is more challenging, as there were around 6.5 million active social media users in Ethiopia by January 2022. While the number of social media users in Ethiopia increased by 500 thousand (+8.1%) between 2020 and 2022, Statista.com projected that the number of social media users in Ethiopia would reach 48.59 million by 2025. Online hate speech in Ethiopia has exponentially risen as of 2018 and is linked to the political instability characterized by conflicts among competing ethnic factions over resources, power,

and narrations and weaker law-enforcing practice (Ayalew, 2020; European Institute of Peace, 2021; Gesses, 2020). Critiques also contend that social media in Ethiopia is ethnicized in a way that polarizes the people (Skjerdal & Mulatu, 2021).

Needless to say, the political turmoil the country has been going through over the past five and six years has caught the attention of various stakeholders including international observers. For example, while the Global Peace Index (GPI) ranks Ethiopia 151th out of the 163 countries included in the analysis, with a 2.87 score on a 1–5 scale, which is judged ‘low in peace’ (2023), in the World Justice Project's (WJP) annual Rule of Law Index, Ethiopia is placed 129th out of 142 countries (WJP, 2023). In addition, weak law-enforcing agents, such as the inability to properly monitor perpetrators and social media platforms and failure to monitor harmful content have put Ethiopians at a higher risk. During the two year deadly conflict in the north, between Ethiopian Federal government and TPLF, for instance, Meta has admitted it failed to remove harmful contents from its platform, leaving Ethiopia at risk of further genocide-level civil war (Meta, 2022; The Guardian, 2022).

The rise of hate speech on social media threaten both social media users as well as offline communities, reaching populations with limited or no internet connection and traveling by text or word of mouth, as a rumor voiced in the marketplace (Mercy Corps, 2021; Stenberg, 2020). In addition, hate speeches on social media not only affect the current generation but also ultimately lay the groundwork for a polarized, fragmented, and dangerous world for the next generation.

Evidences show unless hate speech is nipped in the bud, it is more likely that devastating repercussions including genocide can happen. Pohjonen, for example, noted that an increasingly aggressive and polarized style of communication is one of the telltale signs of escalating conflict (2018). Jews in Germany Nazi, Tutsi of Rwanda, Rohingya Muslims of Burma/ Maynmar, and opponents of the Khmer Rouge regime in Cambodia are among the others who suffered a lot (i.e., deportation, mass rape, and genocide) due to hate speech (Hinton, 2005; Lee, 2021; Morrock, 2010). There is, hence, an urgent call to counteract online hate speech before we go further in the continuum of destruction. In response to the growing problem, nonetheless, divergent calls are made for it to be embraced, tolerated, targeted for counter-speech, blocked on websites, brought to a civil lawsuit, and criminalized.

Wondering the pervasiveness of the concept, a seasoned expert in hate speech, Andrew Sellars, once said, ‘if there is one point of agreement, it is that the topic is ripe for rigorous

study (2016, p. 1)'. In support of the need to conduct studies on hate speech, the United Nations Strategy and Plan of Action on Hate Speech underscored the need to 'undertake quantitative and qualitative analysis of hate speech trends to identify patterns in a given country and regional and local contexts' (UN, 2020, p. 27). In response to the alarming situation, a few studies on hate speech in Ethiopia have come to light in recent years. Pohjonen (2018), yet, claimed the first ever study on online hate speech in Ethiopia was a pilot project, 'Mapping and Analyzing Hate Speech Online, by Gagliardone and colleagues (2014)'.

Although some local studies on hate speech have emerged in recent times, I strongly believe that the gap in the literature remains a concern to be addressed. The available literature on hate speech is deficient, particularly from a social-psychological perspective wherein its antecedents (e.g., intergroup attitude, stereotypes, and prejudice) and consequences (e.g., violence) have higher relevance to the field (Fisher et. al., 2018). More importantly, there are scant empirical studies that examine the types of ethnic-based online hate speech, the contexts in which they are embedded, the mechanisms involved in the occurrence of hate speech, and the roles social media users play in responding to hate speech and hence deciding its fate such as its propagation. This dissertation is, hence, aimed at exploring the various types of hate speech and the mechanisms involved in its occurrence on social media.

1.2 Statement of the Problem

In recent years, Ethiopians from all walks of life, including political leaders, academics, religious fathers, activists, and ordinary social media users both at home and abroad, have been seen contributing to the unprecedented rise of ethnic-based hate speech, endangering the social fabric, dignity, and peace of the people (European Institute of Peace, 2021; FDRE, 2020; Muluken et al., 2021; Megersa & Sinshaw, 2019).

In my causal observation on social media, for example, I am a witness to the fact that ethnic groups are described as animals, weeds, enemies, settlers, cancer, and invaders. Referring to the hostilities on social media, Gessese compared the Ethiopian media platform to the Radio des Mille Collines (RTLHC), a radio in Rwanda known to have contributed to the 1994 Rwanda genocide (2020), while Ilibagiza, a survivor of the 1994 Rwanda genocide and author of the book "Left to Tell," issued a video message to Ethiopian social media users advising us to desist from spreading hate speech (Shumete, 2020). In addition, on December 10, 2020, Meta, formerly called Facebook, ranked the risk of societal violence in Ethiopia

due to unmanaged hate speech on the platform as "dire", its highest threat warning given only to Ethiopia (Meta, 2020). Below, an excerpt taken from a transcript of one of the ethnic-based hate videos analyzed in this study is presented to illustrate the hate speech scenario during the data collection period:

Over the past ten months, all evil things have been committed against us [REDACTED]. These groups [REDACTED] are monsters that *eat the flesh of human beings*. Our enemies aimed to eradicate us slowly. The heinous act they are doing against us is *the continuation of the legacy of their ancestors!* Killing people, cutting hands, and cutting off breasts is their culture! [Now in Afan Oromo]. They are beasts to be isolated in the jungle! They are numb desensitized to human suffering! ... We used to suspect the truthfulness of the brutality of these people when we heard stories from others because we couldn't testify to the claims. Now we have seen their criminal act in our very eyes! [Now in Amharic] You are patients, and you must be cured of your disease. You don't have even the smaller elements of conscience. *Dogs have better consciences than yours*. You killed ethnic [REDACTED] ate their kidneys, and reported being killed by them. The rape and the burning of the bodies of the killed [REDACTED] *will not be forgotten*. The intention of [REDACTED] is to kill all [REDACTED], steal resources such as the Geez and the [REDACTED] histories, and present them as their own! For [REDACTED] the message I have is that we must prepare curriculum and design museums to remember the atrocities. We need not get slowed.

(italics to emphasize, ethnic names shaded)

The above excerpt, which is framed in the context of the two years deadly conflict in the north, possibly provides a glimpse of the nature of hate speech, motives, narratives, self-appointment, stereotypes, anger, the commitment part of hatred, and many more issues worth attending to. Nonetheless, this is just the tip of the iceberg. It appears, hence, that there are ample alarming clocks louder enough to remind us that it is high time for us to fight back against hate speech.

I contend that the rise of online hate speech in Ethiopia is worrisome due to the following reasons: First, the actual number of active social media users and the penetration rate of internet and social media usage in Ethiopia are high. For example, while there were around

21 million internet users with 16.7% internet penetration, more than 6.5 million active social media users are expected to exist in Ethiopia, as of January 2023 (DataReportal, 2023; Statista, 2023). Nonetheless, it is reported that the majority of social media users in Ethiopia are naïve and gullible users (Kinfu, 2017), who can easily be tricked and trapped in the circle of hate speech and disinformation. The rise of hate speech on social media is also partly accounted for by the anonymity it gives to users (Joinson, 2001), and the resulting deindividuation- the loss of one's sense of individuality and personal responsibility (Kokil et al, 2022).

In addition, under certain conditions online hate speeches signal offline conflict (Article-19, 2015). This is perhaps due to the use of social media platforms for sharing hot buttons. For example, a recent study shows that the online hate speech content in Ethiopia during the political reform reflects the country's offline ethnic tensions (Muluken et al, 2021).

Second, ethnicity, as a form of identity, has in the contemporary world a singular capacity for social mobilization as well as psychological properties and discursive resources to decant into violence-although it is not inherently conflictual (Malešević, 2004; Young, 2008). In the same vein, various researchers agree that ethnic tensions appear particularly susceptible to the weaponization of social media (Benesch, 2014; Bossetta, 2018; Mercy Corps, 2021; Singer & Brookings, 2018). Indeed, genocidal acts in Rwanda, Germany-Nazi, Myanmar, South Sudan, and others are caused by ethnic or racial differences (Anstey et al, 2012; Kressel, 2002; Morrock, 2010). Parallel to this, while hate speech can be based on various protected characteristics, the dominant hate speech observed in Ethiopia is that of ethnic-based (Gessese, 2020; Skjerdal & Mulatu, 2021). In support of this, a study by Mulugeta (2019) on 'Mapping Online Hate Speech among Ethiopians: The Case of Facebook, Twitter, and YouTube', shows ethnic-based hate speech counts 81.5%. In this sense, Ethiopian politics, which centers on ethnic identity, is more likely to contribute to the rise of hate speech. A study by Skjerdal and Mulatu, for example, shows that Ethiopian media are used for political agendas favoring a particular ethnic group and hence contribute to polarization and the rise of hate speech (2021).

Based on the idea that social media platforms provide space for discourses that are harmful to specific groups of people in specific contexts (Mondal, 2017), and based on my observations and a few local studies, I argue that not all ethnic groups in Ethiopia are equally distributed or involved in the hate speech saga. Rather, three ethnic groups are mainly implicated in the

hate speech saga, both as objects and subjects: ethnic Amhara, ethnic Oromo, and ethnic Tigrayans (Ayalew, 2020; Gessese, 2020). Consequently, it is imperative that we pay attention to the issue of ethnic-based hate speech on the internet focusing on these three ethnic groups.

I contend that the forces of hate speech and its associates gain strength if left unchallenged. Even more, unaddressed hate speech could attract more individuals to its whirl, as seen in the following comment, which I came across on Facebook, ‘የተኛ ሰው አለ ትቀሰቀሳለህ የሚባለው እውነት ነው በፖለቲካ እና በሃይማኖት ማውራትም አልወድም የሚያወራኝንም ሰው አልሰማውም ነበር ዘረኛ የሆነም አይመቸኝም ሳላስበው እየቀሰቀሳችሁኝ ነው’, which is translated as “The saying ‘slow down, you will otherwise walk up those who are sleeping’ is correct. I am not particularly interested in discussing politics or religious issues, nor do I want to listen to such people, because I’m not racist and don’t want to associate with racists. **But you’re slowly awaking me up from my sleep**” (bold to emphasize). My observation like this and the excerpt on page six shows me there are routes through which individuals are drawn into a hate speech circle. Nonetheless, as to my knowledge, no previous study explicitly explored the routes to hate speech such as how people engage in and drawn into hate speech circles.

Contrary to the rise of hate speech and the accompanying tension, I contend that it has received limited attention for empirical investigation in the context of the current Ethiopia. Parallel to the recommendation by UN (2020) and Sellars (2016) regarding the need to study ethnic-based online hate speech, this problem in Ethiopia deserves a rigorous study that can shed light on its nature and mechanisms involved. Even though, there are few emerging local studies on hate speech, most are from journalism (Hagos, 2018; Skjerdal & Mulatu, 2021), law (Ayalew, 2020; Kinfu, 2017; Tewodros, 2019), and computer science fields and that they check its occurrence, measure its prevalence, examine the hate speech law (i.e., HSDSPPE), or how ethnic nationalists abuse the media. While these local studies have touched on the problem from their perspective, I argue that there is a dearth of literature in the following areas:

First, the available local studies on hate speech use this generic name ‘hate speech’ and never identify its possible manifestation (typology). Yet, as hate speech is the language of hatred (Anstey et al., 2012), and a process and outcome of enemy-making (Oppenheimer, 2006; Stein, 2016), exploring types of ethnic-based hate speech and their context could help us understand how parties involved perceive each other, the nature of hate speech expressions,

as well as the contexts they are embedded. Gorka, for example, argued that when hate speech is directed at an ethnic group, some people begin to believe that the language reflects the thoughts and social attitudes of that group (2019). Such exploration would also give us evidence that shed light on a useful tool to countering hate speech and its associates. I also argue that identifying typologies of hate speech can help Ethiopian legal experts to reconsider the definition of hate speech, which is fraught with criticism for being vague (Ayalew, 2022). Equally important to types of hate speech content is, nonetheless, the context in which they are embedded. While no previous local study showed an explicit interest in identifying contexts of hate speech, I strongly believe that hate speech cannot happen in a vacuum. Studying contexts of hate speech is of paramount importance for several reasons, including understanding intent, assessing the severity of the speech (Benesch, 2014), understanding legality, and tailoring appropriate responses that can tackle not only the hate speech but also the context.

Second, hate speech on social media attracts both like-minded and opposing factions, how users respond to hate speech posts would determine the fate of the post either to fuel further hatred or to make it less dangerous. Indeed, it stands to reason that how people respond to hate speech will influence whether or not such communications are directed towards peace or conflict and hatred.

In this regard, as commenting is an intense activity for many users, who spend many hours daily on it, the type of comments users provides to hate posts have impact on the social media ecology. Studying reaction of users is also helpful as hate speech is self-fulfilling and self-reinforcing. It is self-reinforcing as individuals may first provoke others on the belief that their targets hate them, yet only to attract hate speeches from them and hence witness what they expected first. That is to mean being the subject of hate become real by virtue of expecting others to hate them and acting accordingly. Hate speech also feeds on and strengthens itself. As such, studying responses given to hate speech posts also gives us insight into how hate speech spreads.

Previous researches on the online behavior of users have produced several typologies. For example, Çiçek and Erdogmuş (2013) identified five types of users (e.g., inactives, sporadics, entertainment users, debaters, and advanced users). Kim (2018) identified four types of users (e.g., impression management type, lurker type, enjoyer and relationship focus, and social value orientation type), and Krithika and Kumar (2018) classified users into four categories: socializing, expressing, recreation, and information. In the context of crises communication,

Mirbabaie and Zapatka (2017) identified four typologies of users: information providers, starters, amplifiers and transmitters. Concerning those who disturb others on social media, nonetheless, the term trolling, spamming, cyberbullying, and hate mongers are used in varying contexts such as disinformation (Zannettou et al., 2019), provocation (Mkono, 2015), and harassing and insulting (Cheng et al., 2017). Hate mongers on social media can be motivated by a wide range of factors including prejudice, traits, fun, attention-seeking, or revenge (Lampropoulos, et al., 2022; Shachaf & Hara, 2010) as well as ideology and political motives (Yanal, 2022).

I believe that the available studies on users' online behavior vary across methods (e.g., single platform vs. mix of platforms) and contexts (e.g., political, learning, advertisement, etc.), and most focus on reasons, frequency, and modal activities. Nonetheless, to my best knowledge, there is a dearth of empirical studies regarding how social media users respond to online hate speech, especially in the context of the current Ethiopia, where identity politics is almost at the center of the rise of hate speech. Nonetheless, users' online behavior in the context of conflict scenario is little researched in Ethiopia. Therefore, the second objective of this dissertation is to explore the various online behaviors of social media users in reacting to online hate speeches.

Third, even if hate speech is supposed to be employed by hate mongers to frame ideologies and mobilize resources (Ayalew, 2020), how specifically they manage to trap followers remains a grey area that must be explored. Equally important is to explore the possible springboards from which hate speech is sourced. In support of this, the United Nations strategy and plan of action on hate speech recommend detecting and addressing structural, historical, situational, and individual-level factors contributing to group animosity (UN, 2020). To my best knowledge, nonetheless, no previous study examined mechanisms that are responsible for the occurrence of hate speech, and generated explanatory concepts that can help in describing and explaining ethnic-based hate speech occurrence.

1.3 Research Questions

Based on gaps identified and the priority given, this dissertation is aimed at providing answer for the following questions:

- What are the types of ethnic-based hate speeches circulated on social media?
- How do social media users react to ethnic-based hate speech posts?(typology of users with regard to responding to ethnic-based hate speech posts)
- How do ethnic-based hate speeches occur on social media?(mechanisms involved in the occurrence of ethnic-based hate speeches)

1.4 Significance of the Study

Online hate speech has real consequences for individuals, groups, and societies (UN, 2020 Waldon, 2014), which makes it an important subject for study. It is worth noting that studying hate speech should be approached responsibly and ethically, with a focus on raising awareness, promoting understanding, and fostering positive change. In this regard, I expect that the outcome of this dissertation could contribute in the following ways: First, hate speech is an area with limited social-psychological studies and literature to refer to. Therefore, the findings obtained, such as the types of hate speeches, the context, and the mechanisms involved in the occurrence of hate speech, as well as the roles social media users play in responding to hate speech, help us enrich our knowledge base regarding ethnic-based hate speech. In addition to describing and explaining hate speech and its occurrence in Ethiopia, this study can also be generalizable to other settings.

Second, this study will provide a foundation for future research in this field. Knowledge about hate speech can also empower individuals who are targeted by providing them with resources and support networks to combat the negative effects of hate speech. Recognizing the types of hate speech enables individuals to recognize and report instances of hate speech. By studying hate speech, society can work towards dismantling systemic prejudices and processes of enemy making. A comprehensive understanding of the types of hate speech can help foster a more inclusive society and promote social justice for all. Moreover, studying how social media users respond to hate speech posts provides valuable insights for combating hate speech, fostering digital activism, informing platform policies, and empowering individuals to actively contribute to a more respectful and tolerant online community.

Third, the outcome of this dissertation can help reduce violence and ensure the protection of human rights. For example, the types of hate speech and the contexts in which they embed help understand the mutual perception and the intergroup hatred between groups. This finding can be used to make informed interventions in a way that can restrain negative perceptions and help manage the contexts. The study can also help promote dialogue and understanding between different ethnic communities in Ethiopia by shedding light on the harmful impact of online hate speech and how it is being propagated. This can contribute to peacebuilding efforts and social cohesion.

Fourth, understanding the causes of ethnic-based hate speech can help the government and non-governmental organizations to formulate policies and initiatives that address this issue effectively. Legal authorities, policymakers, and other practitioners who are interested in

fighting back against hate speech benefit from the findings of this study. For example, the knowledge generated can be used to formulate legal frameworks and directives as well as to mobilize people against hate speech. The study can support lawmakers and policymakers in creating a robust legal framework to combat online hate speech. It can provide evidence-based input to draft regulations that balance free speech with the need to protect individuals and communities from harm.

Last but not least, companies, organizations, and projects having an interest in countering hate speech or those who are working on peacebuilding (such as the Ministry of Peace and media platform companies) are expected to be beneficiaries of the findings of this study. For Social Media companies: for example, it allows for improved monitoring of social media platforms where hate speech is prevalent, potentially leading to better moderation and removal of such harmful content. It can guide tech companies in improving their policies to control hate speech on their platforms, which can have global implications, especially in regions where ethnic tensions are prevalent. Projects aimed at enhancing digital /media literacy, instilling responsible social media use, and enhancing intergroup/diversity tolerance to counter negative ethnic stereotypes and hate speech can also use the findings of this study.

1.5 Scope of the Study

This paper doesn't intend to include all media platforms. Rather, a few salient online social media platforms—YouTube, Facebook, and Twitter—are considered. That means the various other social media platforms except the three, print media, radio, and television broadcasts are not included in the study. Second, the contents included in this study are those that were circulating in the last three years during the time of data collection. Therefore, older hate speeches couldn't be part of this study unless they resurfaced as new developments during the time of data collection. Third, both hate and hate speech can occur at various levels, such as individual, group/interpersonal, intimate relationships, and national levels. Yet, this study focuses on the intergroup level of hate speech, particularly in relation to inter-ethnic relations in Ethiopia. Hate speech can also be studied from various angles, such as race, religion, sex, sexual orientation, ideological stance, etc. In this study, nonetheless, only ethnic-based hate speech is focused. In this study, I include only hate speeches based on ethnicity, and irrespective of its intentionality.

While studying hate speech-related fake news, both disinformation (false information with the intention to cause harm) and misinformation (false information without apparent intention

to cause harm) are taken under the umbrella of ‘fake news’. In addition, the languages considered in this study are Amharic and English. Last but not least, as the time in which the data are to be collected is when the political stake is high in the country (such as the internal displacement, killings, and the conflict in the North), any reference to this study should take this time frame into account.

1.6. Definition of Key Terms

- I. **Hate Speech:** in this study hate speech refers to words of incitement and hatred...that advocates, threatens, or encourages violent acts or a climate of prejudice and intolerance (Gagliardone et. al., 2014, P.5), or expressions that are degrading, harassing or stigmatizing which affects a group’s dignity, reputation and status in society by means of linguistic and visual effects based on protected characteristics, ethnicity in this study (Ørstavik, 2015).
- II. **Ethnicity** (ethnic group): is understood in this paper as a concept referring to a culture sharing group and its practices, values, rituals, and beliefs and its way of life as reflected in shared language, traditions, material culture (e.g., clothing and cuisine), and cultural products such as music and art (Malešević, 2004).
- III. **Elites:** in this paper it implies community leaders, academicians, political leaders, activists, artists, etc., which are identified as representing their respective ethnic group or claim to be so.
- IV. **Social Media:** used in this paper to mean “a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0 that allow the creation, post/share, and exchange of User Generated Content” (Kaplan & Haenlein, 2010, p. 61).
- V. **Mechanisms:** the processes or the means through which hate speeches occur on social media, and hence is able to explain how hate speeches occurs and propagated. More specifically, mechanism for hate speech is a constellation of organized entities and activities which regularly bring about hate speech.
- VI. **Typology**, also called type here, is a composite measure that involves nominal classification of observations in terms of their attributes. It is a set of categories used for classifying entities or grouping of observations on the bases of their similarity.

Chapter Two

Literature Review

This section presents an overview of relevant literature from different fields of study that, in my opinion, would help us better comprehend ethnic-based hate speech on social media. It starts with a discussion of hatred, with the understanding that hate speech refers to specific expressions of hatred directed at specific groups in specific contexts (Post, 2009). Second, a discussion of online hate speech is presented, including its various meanings, causes, and consequences. Given the lack of theories on hate speech, I used three larger bodies of literature from social psychology as the foundation for our investigation: stereotypes, social identity theory, and a few selected social influence theories. Then, the section presents a discussion of social media users' online behavior, followed by media use theories. Finally, the section discusses available hate speech countermeasures.

2.1 Concept of Hate: an intergroup aspect

Throughout human history, hatred has caused significant harm to underprivileged and discriminated groups, prompting questions about why people hate others, how they exhibit hatred, and how to address these pressing social issues in our communities (Allport, 1954; Halperin et al., 2009; Ruse, 2022; Sternberg, 2020). Even if hate has attracted scholars from philosophy, religion, social psychology, sociology, political science, law and others, the literature on hatred in psychology is limited by its therapeutic aims and focus on prejudicial hatred and pathological cases (Navarro, 2013). Echoing this, Harrington stated the following

Hatred has not typically been a topic of research in the field of social psychology, although several components which embody hatred have been studied extensively in this field. Social psychologists have traditionally considered prejudice, stereotyping, discrimination and intergroup aggression to be highly important and socially relevant topics for research, and thousands of studies by social psychologists have examined these and other issues related to hatred (2004, P. 1)

Hate is a complex concept with variations in origin and manifestations, leading to a definitional limbo in the literature—a failure to have a comprehensive single definition. Because of its multifaceted nature, attempting to have a single definition is seen as a wrong intent because it limits our understanding of its richness, including types and components

(Fisher et al., 2018; Royzman et al., 2004). For example, hate has been seen as intense and irrational emotion (Gaylin, 2003), as generalized anger and generalized negative evaluation (Ben-Ze'ev, 2000), a motive to devalue others (Rempel & Burris, 2005), a mental abnormality, emotional attitude ²(Ekman, 1992), and set of negative attitudes, a cognitive mechanism, and combination of different emotions (Opatow & McClelland, 2007). Hate, therefore, is conceptualized as an emotion, an attitude, a cognitive mechanism, or perhaps a combination of them. This implies that we might be referring to different things while using this illusive term. The variation in these conceptualizations is worth attending. In relation to inferring hate from the speech, it is likely that we infer any of these conceptualizations. For example, if we see hate as an emotion (like anger and fear, or their combination), it implies it is intense but short-lived, while taking it as an attitude and cognitive mechanism implies it is enduring in its intent and resistant to change.

Different conceptualizations, however, concur that hate is a powerful, intense, persistent, and destructive emotional experience that aims to physically, socially, or symbolically harm or eliminate its targets (Fischer et al., 2018; Opatow & McClelland, 2007; Sternberg, 2003). This is perhaps the reason why Ekman (1992) preferred to use terms like emotional attitude to refer to it. Hate often is viewed as a transaction between two parties: perpetrators and targets; more importantly, a sustained transaction. Hate is based on perceptions of a stable, negative disposition of persons or groups (Staub, 2005). We hate persons and groups more because of who they are, than because of what they do. Individuals may report hate if appraising an event as contradicting their interests (relevant to all negative emotions), unjustified and unfair behavior of others (i.e., anger), seeing others as morally inferior (i.e., contempt), or seeing others as morally nauseating (disgust), especially when the action in question is perceived as intentional (Halperin et al., 2009). The hater sees the object of their hatred as bad, immoral, dangerous, or all of this together (Martínez et al., 2022; Staub, 1992). Hate is first and foremost used as shorthand for a range of specified -isms and phobias (Brudholm & Johnsen, 2018). Below, a framework by Strenberg is presented which I think is comprehensive and appealing to this paper.\

² A term he reserved for affective phenomena that are more sustained and typically involve more than one emotion

2.2 Duplex Theory of Hate

Robert J. Sternberg (2003, 2005), a well-known psychologist, has famously discussed hate and developed a duplex theory of hate consisting of two component theories: the triangular theory of hate and the theory of hate as a story, which are presented below.

2.2.1 Triangular Theory of Hate

According to this theory, hate is comprised of three dimensions: the negation of intimacy (affective), passion (motivation), and commitment (cognition of devaluation). The affective component is characterized by feelings of repulsion and disgust towards the object of hatred triggered by the characteristics of the hated target as well as by propaganda or narratives. Individuals with a high score on this component either avoid or prevent their object of hatred from approaching or entering a contact setting such as friendship circle, marriage, or neighborhood.

The passion (motivation) component of hate manifests as intense anger or fear in response to a threat, and thus fight-or-flight is the reaction pattern in the presence of the hated object. According to Sternberg (2003), this component of hatred is typically rapid in its emergence and demise. The third component, commitment, entails devaluing and reducing an individual's or group's human characteristics, viewing the target as barely human or subhuman, profaning or degrading the target's reputation (Navarro, 2013; Sternberg, 2005). The goal of those who incite this type of hatred is to change the thought processes of the preferred population, particularly their in-groups, so that its members perceive the targeted group(s) as devalued. Brainwashing programs as well as entertainment mechanisms like jocks and cartoons, could be used to devalue the target.

The above descriptions of the three components of hate entail that disgust, fear, anger, and contempt are emotional states accompanying the concept of hate. Each of the components of hate per se gives rise to three different types of hate named cool hate, cold hate, and hot hate; while their combination give rise to four additional types of hate: seething hate, boiling hate, simmering hate, and burning hate, totaling seven in number (Sternberg, 2005). While describing each of the seven types of hate is beyond the intent of this review, the pictorial representation of the typologies is presented below:

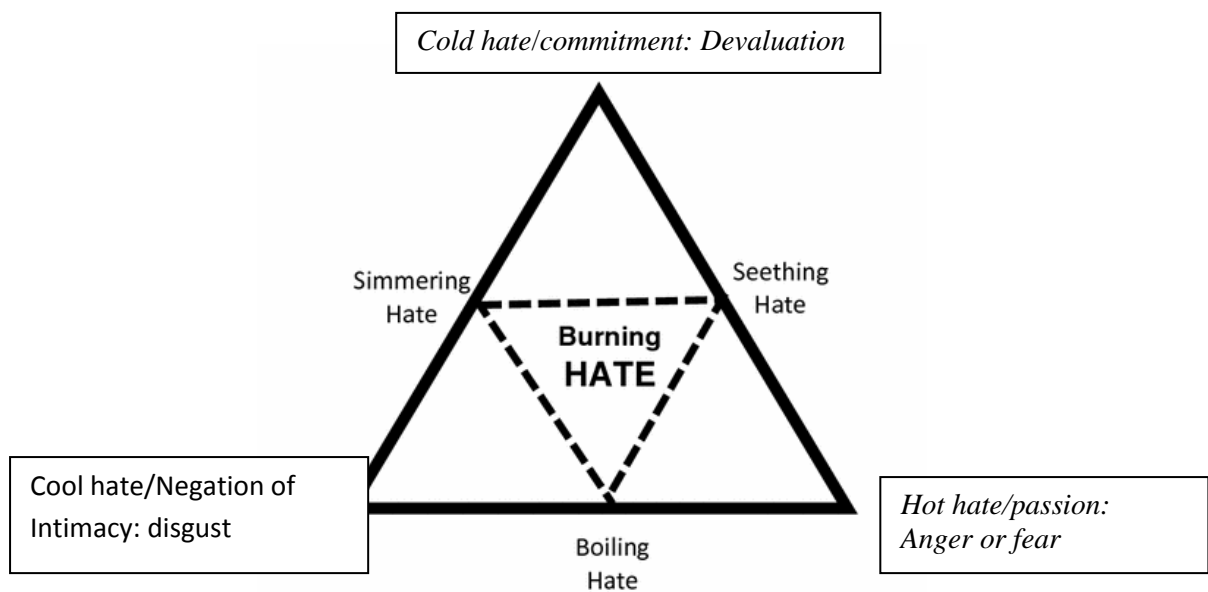


Figure 1 Sternberg's Triangle of Hate
 (Source: Adapted from Sternberg's Triangle of Hate, 2003)

Considering hate speech as a language of hatred (Anstey et al., 2012; Brindle, 2016; Kennedy et al., 2023), it is therefore reasonable to assume hate speech expressions can include the three components of hate. It is also likely that hate speeches is about the expressions of the negative emotions (or emotional attitudes) like fear, anger, disgust, and contempt one holds against its target. This triangular theory of hate also entails the variability of the hate speech expressions as each of the components combines with another and manifests various types of hate speeches.

2.2.2 Hate as a Story

Sternberg argued a set of stories exist that are directed towards the hated object. For example, stories are developed that portray the hated group or person as an enemy of democracy or people, a power-hungry villain, a thief that steals property from others, a vicious and backstabbing enemy, a subtle infiltrator, a detested animal such as a cockroach or rat, and so on (2005). According to Sternberg, different stories are generated in each component of hate. As most people are either unaware of or only vaguely aware of their own stories, they may find themselves hating individuals or groups without fully understanding the psychological reasons for their hatred (Sternberg, 2020).

Intergroup-emotions theory, relative deprivation theory, and social dominance theories are effective explanations for how such stories contribute to the development and justification of

hatred between groups. Social-dominance theory of Sidanius and Pratto (1999), for example, assume that powerful groups retain power by virtue of myths that help to legitimize their power. Real, symbolic, and even completely imagined threats to diverse groups, governance systems, and established power hierarchies have the potential to spark hate stories which in turn, accelerate the development and transmission of hate.

In short, hate is a compound negative emotional construct described more particularly as a long-term set of negative attitudes, motivations, emotions, and dispositions against a target (Sternberg, 2005) and is filled with pride, humiliation, animosity, hostility or fear, and/or anger, a worldview, or an ideology (Opatow & McClelland, 2007) that tempts one to reject others in a generalized and totalistic fashion (Ben-Zeev, 1992). Concerning whether hate speech manifests hatred towards its target, there seems to be competing or alternative views. Dreu and colleagues (2016), for example, argued that verbal arguments during intergroup conflicts, in-group defence is stronger and better coordinated than out-group aggression. Nonetheless, Bizumic and Duckitt (2007) explores different types of group self-centeredness and their relationship with negativity towards specific outgroups, showing that ethnocentrism is positively correlated and associated with negativity towards relevant outgroups. Ethnocentrism is regarding one's own social group as the center of all things, as seen in the tendency to judge one's group as superior to other groups (APA, 2015).

2.3 Causes of Intergroup Hatred- *social psychological viewpoints*

The variability in hatefulness amongst human beings is rich, textured, and culturally patterned, and it more likely stems from the confluence of biological, emotional, psychodynamic, cognitive, and social forces (Osborne and Frost, 2004). In his recent book on '*Perspectives on Hate: How It Originates, Develops, Manifests, and Spreads*', Sternberg (2020) identified a few enabling conditions that play greater role in the development and transmission of hate, including fear, obedience to authority, trust, sense of belonging to a valued group, amplification of arousal, and modeling (FLOTSAM for short). Below, a brief description of what can be called 'traditional view of psychology' is presented followed by social psychological causes of intergroup hatred.

2.3.1 The Psychodynamic View

Vicious haters can morally disengage from their actions and make excuses for the hatred they feel or the suffering they intentionally cause. According to Freud's theory of the death instinct and defense mechanisms, hatred is frequently linked to disturbances in the psychic structure.

he also assumes hatred is an important pathological condition in its own right (Freud, 1901, in Fitzgibbons, 1986). For example, he argued that an unconscious and unacceptable "wish" or impulse can become a conscious fear, fueling hatred for the repressed component of this dynamic. In essence, hate is an emotion that masks personal insecurities (Oppenheimer, 2006). This perspective holds that humans have a psychological obligation to find ways to express disturbing thoughts. Similarly, individuals receive psychic benefits if they despise the people they intend to rob or hurt.

2.3.2 The Self: Identity and Intergroup Hatred

Identity, which is individual's sense of self can be of three types: personal, collective, and relational identities, each activated in different contexts with different implications (APA, 2015; Dovidio et al., 2005). While personal identity is the perception of oneself as a distinct individual with relatively unique characteristics, relational identities are founded on personalized bonds of attachment within significant relationships (Sedikide & Brewer, 2015). Collective identities, on the other hand, emphasize our connection to others who share similar characteristics and identifications through impersonal bonds, do not necessitate close personal relationships among group members, and distinguish in-group members from members of relevant out-groups (Baumeister & Vohs, 2007).

Social Identity Theory (SIT), one of the most important and widely cited analyses of the role of identity in intergroup relations, examines how our identity as members of a group affects our behavior and perceptions towards relevant others (Turner, 1985). SIT maintains that individuals categorize the social world into 'we' and 'they', identify themselves to 'we', compare their group against relevant outgroups on important issues such as resources and status, derive some portion of their self-concept from it, and develop biased thinking towards each (Turner & Oakes, 1986). In this regard, in his classic work 'The Nature of Prejudice in the chapter 'In-group Formation', Gordon Allport (1954) postulated that in-groups are "psychologically primary," in the sense that familiarity, attachment, and preference for one's in-groups come before the development of attitudes towards specific out-groups. In-group bias refers to a form of favoritism towards one's group, its members, its characteristics, and its products compared to other groups. In-group bias is not just a matter of rational self-interest but may also include more symbolic and emotional benefits to the group (Baumeister & Vohs, 2007).

SIT posits that a portion of one's self-concept is dependent on the importance and relevance placed on the group membership to which an individual belongs (Turner & Oakes, 1986;

Verkuyten & Brug, 2004). The collective self is based on intergroup comparison processes and is associated with preserving the in-group's image (Brewer & Gardner, 1996), which helps individuals boost their self-esteem. According to SIT, individuals seek to achieve and maintain a positive social identity. To achieve this, people discriminate against out-group members relative to in-group members (Turner, 1985). According to Luhtanen and Crocker (1991) and Osborne and Frost (2004), belittling, derogating, or discriminating against other social groups is one of the mechanisms used to boost self-esteem.

There is also a strong tendency for a fusion of identities. Fusion of identity occurs when an individual's personal and social identities become functionally equivalent within a group (Swann et al., 2009). Baumeister and Finkel summarized this by stating that "the most prevalent theme of the group member involves some loss of individuality within the group" (2010, P. 20), and this human tendency mostly ushers in negative consequences such as groupthink and mob violence (Forsyth, 2010). In a study titled 'Deindividuation and anti-normative behavior: A meta-analysis', Postmes and Spears also reported that when we are part of a large crowd, we are more likely to obey the norms of this group—whatever those may be (1998).

According to realistic conflict theory, when two or more groups compete for limited resources (real or perceived scarcity), prejudicial feelings and negative stereotypes of the opposing group can emerge. The theory holds that conflict with and negative attitudes towards other groups can be reduced if both groups share a superordinate goal (Sherif, 1966). Various scholars that wrote on hate speech agree that one's perception of the target of hatred plays a greater role than the objective reality at hand. For instance, hatred could be provoked as a result of a real or perceived offence against one's members, which is perceived to be rooted in the stable, evil character of the hated group or its members (Halperin et al., 2009; Royzman et al., 2004). In this line, hatred is a cover emotion for feelings of pain, hurt, fear, and threat. Hate, therefore, has a unique pattern of appraisals and action tendencies, basically withdrawal/repulsive or annihilation (Fischer et al., 2018). Since we likely hate when we are insecure, it is useful to look at how haters perceive history and how they feel threatened.

As attitude has some psychological functions such as value expression, ego defense/externalization, and social adjustment/identification, hate can be reassuring, self-protective, and helps confirm people's belief in a just world (Fisher et al., 2018). Once we are in a state of hatred, we are more likely to justify any action against our targets.

Two processes, based on SIT and attribution theories, help to increase group self-esteem. The first is the human desire to be a part of a successful group, also known as basking in reflected

glory (BIRG). This entails enhancing one's association with a prestigious group, frequently using the term "we" to refer to the winning scenario (Leuven et al., 2002). The second strategy is Cutting off Reflected Failure (CORF), the severing of associations with others who have failed in the interest of avoiding a negative evaluation by others (Spinda, 2011). CORF appears difficult because individuals may not cut themselves off from their social group if they are intimately connected. Instead, they could justify the failure and blame it on various internal or external factors.

Groups are supposed to reduce uncertainty about who you are and how you and others will behave. Hogg argued that when self-uncertainty is present and/or activated, people are more likely to engage in discriminatory comparisons to others, which results in denigrating others (2001). When group identity is under threat (e.g., the fear of being absorbed or assimilated into some larger group), members tend to become more exclusionary. As a result, examining the security of the members' social identities as well as the mechanisms by which the group serves to benefit its members is critical to preventing intergroup hatred.

Individuals who identify with their group tend to hear more, know more, care more, and develop a positive attitude towards their group and its members than outgroups. However, whether in-groups develop a negative attitude towards their respective out-groups is debatable. Scholars such as Allport (1954) and Brewer (1999) believe that attachment to in-groups does not always result in a negative attitude towards relevant out-groups, whereas Sumner's (1906) original treatment of ethnocentrism considers the importance of a reciprocal or opposite attitude between in-groups and out-groups. Brewer (1999) contends that social categorization causes pro-in-group bias rather than anti-out-group bias, and thus, even when out-group members are perceived positively, in-group members are perceived and treated even more positively. This subtle form of favoritism is not benign, of course, because it may be associated with or manifested in the absence of positive feelings towards those outgroups (Dovidio & Gaertner, 1993).

In line with this domain of finding, evidence from various social psychological studies showed that, compared to in-groups, out-groups are less likely to be helped in ambiguous circumstances (Frey & Gaertner, 1986), more likely to be seen as provoking aggression (Rogers & Prentice-Dunn, 1981), less likely to receive the benefit of the doubt in attributions for negative behaviors and likely to be seen as less deserving of public welfare (Pettigrew & Merteens, 1995)

When social identity is psychologically important, it affects perceptions, attitudes, feelings, and behavior. Hatred is motivated by an individual's or group's immutable characteristics,

ethnic background, or religious identity, and it causes harm (Sellars, 2016). As a result, identity unites and divides people. In this sense, hatred exists where groups of people congregate or where people attempt to determine who is "one of us" and who is "one of them." It is natural for dislike to appear, but it may take the pathological form of hate.

2.3.3 Cognitions in Hatred

It is believed that the way we think or process information about ourselves and our surroundings (i.e., social cognition) is most assuredly germane to understanding the anatomy of hatred (Osborne & Frost, 2004). Social cognition refers to how we interpret and analyze (i.e., reappraise the assigned meaning) and remember (with effort or automatically) information about other people and ourselves (Myers, 2010; Pennington, 2000).

One of the key facts threaded throughout the discussion on the role of cognition in hatred is the essence of schema, which is a cognitive representation of a category, such as a group of people, and is a unified, stable structure that resists change or modifications. Schemas contain general expectations and knowledge of the world and hence facilitate and guide our cognition (i.e., encoding, memory, expectation, and perception and interpretation), emotion and feeling, behavior, and interactions in the social world (Pennington, 2000). In viewing the world through "schema-colored glasses," we subject virtually all the incoming information to varying degrees of distortion, misinterpretation, and invalidation.

The schema we have about a particular social group is the typical characteristics we assign to it (i.e., stereotypes). Hence, hatred could be the result of the cognitive representation of a particular social group. One of the dangers of social categorization is that, when categorized into separate groups, people are "depersonalized," or seen as embodying a relevant group prototype (Jackson & Hogg, 2010). In relation to hatred and negative social relations, for example, this concept implies that 'if I know the group you belong to, I don't need to exert effort to know who you are, and I believe you possess the attribute of your group.' Linking this with social identity theory, prototypes maximize entitativity (the extent to which a group is a distinct entity) and optimize metacontrast, i.e., the extent to which there is similarity within and difference between groups (Hogg & Smith, 2007). Yet, it is also noted that in-group prototypes can change as a function of which out-group you are comparing your group to. In this way, prototypes are context-dependent.

Because of prototype-based perceptions of outgroup members, stereotyping, all members of an outgroup are given similar stable attributes; everyone within a group is the same (Chin, 2004; Fisher et. al., 2018). As prototypes describe and prescribe group-appropriate ways to

think, feel, and behave, individuals are more likely to conform to the group prototype, leaving those who deviate from it to be disliked and rejected (Baumeister & Vohs, 2007; Levine & Hogg, 2010). In a geopolitical context and genocidal incidents, stereotypes tend to derogate and dehumanize people, placing all members of a group, including children, into one homogeneous category (Allport, 1954; Sternberg, 2005).

Hatred and violence stem from seeing oneself under attack and describe many hate crimes that were felt to be a response to bad actions on the victim's side (Marcus, 2012). While collective victimhood keeps the memory of hate alive across generations, it may also direct the appraisal of future events. In this sense, self-preservation motivates attention towards cues that tell one that one is under attack and demand to take protective measures (Beck, 2001). If there is perceived or real competition over scarce resources in a zero-sum game template, degradation, hatred, and name-calling intensify and become part of the cultural tapestry.

Errors individuals commit during social cognition, which emanates from schematic thinking, is also worth attending. Using the words of Fiske and Taylor (1984), we are all cognitive misers—that is, we minimize the time and energy required for information processing by making social reality as simple as possible (i.e., heuristic). Heuristics, which relies too much on schemas, help us process information quickly and economically and facilitate memory recall but mostly result in a number of cognitive distortions responsible for negative emotions and feelings such as hatred and prejudice (Stangor & McMillan, 1992). In the book 'prisoners of hate: The Cognitive Basis of Anger, Hostility, and Violence' Aron T. Beck Explains how the offenders are locked into distorted belief systems that control their behavior and shows how the same distortions in thinking occur in a rampaging mob as in an enraged spouse (2010).

A summary of the works of scholars such as Navarro (2013), Beck (2001), and Shiraev and Levy (2013) gives us the following cognitive errors that have high relevance to the study of hatred and hate speech: Schemas enable us to "fill in" the gaps in knowledge by using "best guesses" and "default options" based on our preconceptions and prior experiences. This leads to availability bias, a cognitive error that happens when we rely too much on information that is readily available to our minds, such as firsthand accounts, dramatic tales, remarkable incidents, and unusual events, which can skew or distort our judgments. In support of this, evidence shows that hatred usually makes us remember well-publicized misdeeds of the outgroup and makes us believe they always behave so. In addition, some schemas contain a strong affective component, so that when they are activated, the associated emotion is cued (Fiske & Pavelchak, 1986). In what is called belief perseverance, for example, once we

become personally invested in a belief, we then tightly cling to it, even facing information that contradicts our belief (Fiske & Dyer, 1985). One of the perplexing cognitive errors we are all likely to engage in is acutely feeling the reality of something without any basis. This is the essence of one of the cognitive errors named emotional reasoning—a cognitive process by which an individual concludes that their emotional reaction proves something is true despite contrary empirical evidence.

In a group context, people feel that others hate them because they feel hated inside. Individuals with such emotional reasoning are more likely to engage in accusing others of the very emotions and feelings they are experiencing inside. Individuals also commit the Barnum Effect- which is the tendency to believe that vague predictions or general personality descriptions have specific applications to oneself or particular groups.

The fifth cognitive error implies that we are likely to explain the behavior of others as resulting predominantly from their personalities, whereas we often minimize (or even ignore) the importance of the particular context or situation (Mayers, 2010). With regard to hatred, for example, any misfortune that happens to the outgroups will be understood as if it is caused by their inherent characteristics or internal traits. Representativeness heuristic or bias is when an individual's action is taken as if it represents the behavior of a group the individual is a part. In assimilation bias: we view the world through “schema-colored glasses,” and we subject virtually all the incoming information to varying degrees of distortion, misinterpretation, and invalidation.

In a related manner, haters are also seen as engaged in the just world belief, which according to Lerner (1980), individuals are motivated to perceive the world as a just place, in which people have what they deserve and deserve what they have. In the just world belief, individuals engage in biased inference in order to preserve their Belief in a Just World, thus attributing the disproportionate victimization of a group to that group's negative characteristics. This leads individuals to feel apathy to the misfortunes experienced by individuals, particularly members of the outgroup (Dharmapala, et al., 2008).

Considering correlation as causation is still another cognitive distortion we are all vulnerable to. Correlation as causation is a belief or expectation that helps to bring about its fulfillment. Haters, hence, take the opportunity of two correlated things to augment their negative feelings towards outgroups, believing that one is the cause and the other is the effect. Insults, accusations, and labeling can easily cause this kind of cognitive error. Last is confirmation bias, which is the tendency to interpret the world to fit our existing beliefs, ignoring or neglecting counterevidence. Individuals are likely to examine relevant empirical evidence in

a biased manner, to accept "confirming" evidence at face value while subjecting "disconfirming" evidence to critical evaluation (Lord et al., 1979). When we quarrel with our friends, we might recall their past mistakes vividly but are reluctant to acknowledge our own.

2.3.4 Sociopolitical and Historical Contexts of Hatred

It is important to note that the study of hate cannot be fully understood without considering the sociopolitical context in which it exists (Opatow & McClelland, 2007). It is more likely that our social milieu influences who we are and how much we hate. Hatred is formed when people believe that outgroups are responsible for past crimes. More importantly, group hatred develops because politicians supply stories of past atrocities to their audiences to get support from those who listen to them (Bar-Tal et al., 2014). As elegantly described in SIT, the story of self begins to be constructed within the context of a larger story—a shared story of who we are. When a child's development occurs within the narrative context of an exclusivist story coupled with hate, the likelihood of a self-story patterned by hate increases exponentially. In support of the political role of hatred, Yanal also assumes hatred as an ideology (2022).

The root of conflict is often generations in the past. By the same token, an important cause of hatred is the issue of history. Ethnic groups often make historical claims based on a one-sided interpretation of events that are taken as facts. History not only defines the group, but it is also the heart of the conflict (Nadler et al., 2008). Narratives are also highly relevant in understanding intergroup hatred. Each identity group is constituted through the narratives it tells about itself and others, as well as the stories told about it by others (Ross, 2001). It could be an origin myth, or it could function as a narrative that describes the current challenges for "our" group.

Hatred is the outcome of a political market in which self-interested political entrepreneurs interact with everyday citizens. Politicians supply hatred if it complements their policies or strategies (Bar-Tal et al., 2014; Glaeser, 2002). However, given the nature of the "public" and its issues, political narratives can either contribute to the development of the community itself, or they can exaggerate existing fractures within a community and radicalize groups (Graef et al, 2018). By same token, Covell contend that contemporary attitudes toward political leaders and media exposure influence attitudes toward war (1996).

As Sam Keen argues (1986), many children are taught stories that denigrate specific others. The stories may explicitly assign derogatory labels to others like savages, heathen, godless, vermin-or, in more subtle ways, cast others in antagonist roles. This process of dehumanization, whether explicit or implicit, may even be conveyed in jokes, in the literature

of comedy; even then, the toxic message is conveyed— if not heightened (Frost et al., 1992). As we construct and tell our life stories, we may be motivated by peers, authority figures, and other social influences to define ourselves and to get on with our purpose (Habermas & Bluck, 2000). Individuals who fear the out-group fail to differentiate differences among all members of the out-group and view all of them as dangerous and evil. This fear of specific groups becomes chronic and is adopted into the group's history (Covell, 1996) and invades every pore of society (politics, literature, school books, media, and religion (Bar-Tal, 1993). One main source of hatred and violence is threatened ego, particularly when it consists of favorable self-appraisals that may be inflated or ill-founded and that are confronted with an external evaluation that disputes them (Baumeister et al., 1996). People may feel threatened if their positive self-concept is attacked by outgroups. This, in turn, can trigger hatefulness towards that individual or the group one belongs to. This hatefulness will increase if it occurs in groups. That appears to be one reason why haters rarely hate alone; rather, they find others with similar attitudes or views and join or form a group. The other political cause is a lack of rule of law. In places with limited rule of law, such as sub-Saharan Africa, they are open to ethnic hatred (Easterly & Levine, 1997). In conclusion, historical causes of hatred are typical manifestations of recurrent conflicts such as those seen between Israel and Palestine (Bar-Tal et al., 2014). Summarizing all these, in support of this Opatow and McClelland (2007) contend that hatred is fueled by pride, humiliation, or fear, and/or anger, worldview, or an ideology.

2.4 Hate Speech: The Language of Hatred

2.4.1 Conceptualizations of Hate Speech

Hate speech is a well-known term in today's digital age, used primarily as a catch-all for a variety of meanings. Despite its prevalence and virulence, there is no single, universal definition for this hate speech. Rather, its meaning is the subject of heated debate among academics, practitioners, legal drafters, and policymakers (Nadim & Fladmoe, 2016; Sellars, 2016; UNESCO, 2015). For example, countries with hate speech laws vary in the way they conceptualize and measure hate speech expressions, with varying standards for defining and limiting 'hate speech'. It is probably among the few pairs of words that evoke a diverse range of feelings, perspectives, and reactions (Article 19, 2015).

A wide range of content may or may not meet the definition of hate speech, depending on the context (Sellars, 2016) and perspectives of the beholder.

The difficulty in adopting a universally accepted definition stems from a variety of interrelated factors. First, many find it difficult to adopt a broad definition of hate speech that encompasses emerging expressions and forms of communication. This is because hate speech could be camouflaged with free speech in many ways and forms (UNESCO, 2023). To epitomize this, Brown (2017) said, ‘the term ‘hate speech’ can be applied to countless permutations of relatively unlimited types of context, speech content, emotions, feelings, or attitudes, speech, and activity. This suggests strongly that we are dealing with multiple meanings’ (P. 3).

Second, considering a message as hate speech is a decision that may incite a tug-of-war between respecting the right to freedom of expression on the one hand and the duty and responsibility of protecting others from any kind of violence and crime on the other hand (Nadim & Fladmoe, 2016; Roussos & Dovidio, 2018). Therefore the values and the commitment individuals and organizations have to these issues will influence both what to be included in the definition of hate speech and the contexts that must be fulfilled to justify contents as such. Third, each country, regional and international organizations, and internet intermediaries such as Meta and YouTube give priority to their own ‘business’ and local realities while defining hate speech (UNESCO, 2015). For example, Germany’s laws restrict expressions that affront “the human dignity of others by insulting, maliciously maligning, or defaming segments of the population,” and New Zealand bans speech with “threatening, abusive, or insulting... words likely to excite hostility against or bring into contempt any group of persons... on the ground of the color, race, or ethnic or national origins of that group of persons” (Waldron, 2012, p. 8).

Definitions have also been adapted over time to address new situations and to accommodate shifts in language, shifting understandings of equality, the harms of discrimination, or developments in technology (Article-19, 2015). Conversely, more subtle language attacking an outgroup, which can be harder for casual observers to identify, may have damaging effects on individuals and group relations. This is especially true in the online sphere, where speech is rapidly changing and can be highly specialized (Gagliardone et al. 2015). The use of code words as stand-ins for racial slurs is also common in online communities, further complicating the definition of hate speech (Duarte et al. 2018; Roussos & Dovidio, 2018). Definitions can also vary greatly depending on whether the creator is seeking a theoretical or academic understanding of the concept or trying to intervene against it.

Below, a few definitions that I think are helpful to appreciate the diversity of opinion on this term are presented. First, I start with a definition from the UN, the Ethiopian government, Meta, and a final one from scholars, which I take as the working definitions for this paper.

UN Strategy and Plan of Action on Hate Speech defines hate speech as: ‘Any kind of communication in speech, writing or behavior, that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are’ (UN, 2019, P. 2.). In its description part of the definition, it also divides hate speech as lawful hate speech (i.e., shouldn’t be criminalized rather has to be protected) and unlawful hate speech (i.e., should be restricted by law). This definition, which is drafted in response to the ‘alarming trends of growing xenophobia, racism and intolerance, violent misogyny, anti-Semitism and anti-Muslim hatred around the world, doesn’t talk about intentionality. The Ethiopian Hate Speech and Disinformation Prevention and Suppression Proclamation No.1185, nonetheless, defines hate speech as ‘speech that deliberately promotes hatred, discrimination or attack against a person or an discernable group of identity, based on ethnicity, religion, race, gender or disability (FDRE, 2020, P. 1).

The definition received a barrel of criticism for lacking adequate clarity and specificity (Ayalew, 2022; 2020; Tewodros, 2019). This definition, for example, requires intentionality of the action which is difficult to provide evidence in most situations. Indeed, intentionality is also implicated in other definitions of hate speech which are meant for legal and political purposes (Corlett & Francescotti, 2002). The definition by FDRE also rules out expressions which are scientific and artistic in nature couldn’t be considered as hate speech. This perhaps creates gap that challenges holding hate mongers accountable for their offense (2020).

In a way that adds a layer of difficulty in detecting hate speech, Meta defines it as ‘a direct attack against people, rather than concepts or institutions, on the basis of what we call protected characteristics: race, ethnicity, national origin, disability, religious affiliation, sexual orientation, sex, gender identity and serious disease. It defines attacks as violent or dehumanizing speech, harmful stereotypes, statements of inferiority, expressions of contempt, disgust or dismissal, cursing and calls for exclusion or segregation (Meta, 2023). Among others, this definition also entails ‘direct attack’ which is both difficult to make sure how directness is assessed as well as it rules out indirect attacks. The definition also states that attacking institutions is possible, which in the context of Ethiopia could be highly related with identity.

Let’s examine how two widely cited seasoned hate speech experts define this complex term: While Andrew Sellars understood hate speech polarizing expressions that vilifies, humiliates

or promotes intolerance and violence against groups of persons by explicit or indirect reference to their protected characteristics (2016), Fortuna and his colleagues (2018) defined it as language that attacks or diminishes, that incites violence or hate against groups, based on specific characteristics, and it can occur with different linguistic styles, even in subtle forms or when humor is used.’

I am sure now that the above five definitions clearly show the diversified ways hate speech is defined and the associated difficulty in ruling out what it is or is not. I argued that irrespective of the variations in the above definitions, many hate speech definitions have commonalities. For example, most agree that hate speech is associated with negative attitude towards the target group, as well as it can lead to the development of animosity, intolerance, and exposes its targets to some kind of violence. In addition, while some include hate speech to target individuals, most definitions assume it targets protected characteristics like sex, race, ethnicity, religion or other collective identities.

The opinion on what constitutes 'hate speech' and when it can be prohibited varies greatly, but it includes disagreement on the degree of focus given to the content and tone of the expression; the degree of focus given to the amount and timing; the intent of the speech; who made the speech; and the imminence of harm caused by the speech (Article-19, 2015). In addition, in deciding whether a speech fulfills the criteria of hate speech or reaches the ‘threshold’, many scholars and legal experts agree that the speaker, the context of the speech, the recipient of the message, the target, and the message itself should be well examined. So, hate speech is not only about the question of what is said; rather, it is also about who said it, for whom, in what context, and ultimately the imminence (i.e., causality) and level of harm it is believed to cause (Benesch, 2014; Council of Europe, 2009; Pohjonen, 2018). Hate speech should not be seen as a universal category but rather as a situated practice that always exists in specific cultural and political contexts and media environments.

Hate speech is a term of art, referring to the particular expressions of hatred against particular groups of people in particular contexts (Post, 2009, p. 123). Most commonly, it is understood to be bias-motivated, hostile, discriminatory (biased or intolerant) or “pejorative” (demeaning slurs and epithets), and malicious language targeted at a person or group because of their actual or perceived innate characteristics (Cohen-Almagor, 2011). In an online environment, hate speech should further be understood as the tactical employment of words, images, and hate symbols (e.g., burning crosses), as well as conspiracy theories, politics, and even pop

culture (Corlett & Francescotti, 2002; Heinze, 2016). Hate speech thus is an identity under attack.

In hate speech, it is common to trace ethnocentric thoughts as well as Barnum Effects in which in-groups intentionally assign positive traits to themselves and implicitly let out-groups feel they are deprived of those traits.

While the word hate speech is a not-easy term by itself, there are also related terms that are conflated with this contentious term, such as offensive speech, aggressive speech, harassment, cyberbullying, dangerous speech (Benesche, 2014), incitement to violence, and so on. Some still prefer to focus on specific terms to refer to hate speech, such as dehumanization and delegitimization (Bar-Tal, 2014). Dehumanization, which is the denial of humanness, for example, is conceptualized in various forms. In the ABC model of dehumanization by Tipler and Ruschel (2014), for example, individuals are portrayed as lacking emotion and feelings, cognition (thinking and holding beliefs), behavior (agency), or a combination of these. This can result in seven types of dehumanization and a wide range of implications as a consequence. If portrayed as wild animals, for example, it implies they behave with emotion unchecked by reason and prescribes strategies of domestication and dominance. If they are automata, it means the emotion or affect agency is denied and could mean they threaten in-group goals (Tipler & Ruschel, 2014). Another model of infrahumanization (Leyens et al., 2000) classifies emotion into primary (shared with other animals) and secondary emotions (supposed to be uniquely human emotions) and assumes individuals and groups can be dehumanized by denying the secondary emotions. In infrahumanization, instances of in-group bias, individuals are not reluctant to attribute human nature, such as primary emotion (which we share with other animals), to outgroups. By the same token, Haslam (2006) advanced the primary and secondary emotion classifications into human nature (human traits that can also be possessed by animals, such as anger and fear) and human uniqueness (traits only humans possess, such as civility, rationality, maturity, shame, love, and hope) and identified two types of dehumanization, mechanistic and animalistic, which happen as the result of denying individuals human nature and human uniqueness, respectively.

In mechanistic dehumanization, individuals are more likely to be treated as objects, lack emotional warmth and depth, and lack agency, while in animalistic dehumanization, targets are treated as some kind of animal—to mean irrational, amoral, lacking in culture, and motivated largely by instinct (Leyens et al., 2007). The mechanistic form of dehumanization involves a sense of horizontal comparison—targets are seen as nonhuman more than

subhuman. Hence, while the mechanistic form of dehumanization implies indifference, animalistic dehumanization implies disgust (Haslam, 2006).

2.4.2 Causes of Hate Speech

The causes of hate speech are myriad and complex. Even though the causes of hate speech are myriad and complex, I categorized them into individual level factors, and narratives and sociopolitical factors.

2.4.2.1 Individual level factors: Evidences show certain personality traits, such as antisocial personality disorder (Livesley, 2001), borderline personality, anger, sadism and low level of agreeableness (Brogaard, 2020; Toegel & Barsoux, 2012), the dark triad of personality traits (narcissism, Machiavellianism and psychopathy) as well as social dominance orientation (Dellagiacoma et al, 2024) are found to be positively correlated with involvement in online hate speech. In same vein, Sigmund Freud noted the existence of pathological hatred, which can be revealed through derogating others. In support of this, a study by Górká on online aggressive behavior, for example, showed half of the hate speeches observed are targeted at random individuals, showing haters randomly offend anyone whom they come across online (2019). As this survey shows, the main motive (7 out of 10 people) for hating is the opportunity to relieve the tension and getting rid of frustration; while another reason for hate speech in cyber space is the desire to hurt and humiliate people (Górká, 2019).

Individual level factor also takes us to social identity. According to SIT, for example, individuals seek to achieve a positive collective self-esteem partly by derogating out-group members (Baumeister & Leary, 1995). By same token, individuals could join hate speech circles due to affiliation motive, which is a desire to fit in and not to be at odds with one's own group (Osborne & Frost, 2004). Aside from personality and social identity, low level of media literacy (Glaeser, 2002), low level of diversity tolerance, and poor emotional intelligence are found to be associated with level of hatred and its expression (Goleman, 1995). In support of this, studies show promoting tolerance reduces hate speeches and related sentiments (Council of Europe, 2021).

2.4.2.2 Narratives and Politics: parts of hate speech are the result of negative intergroup narratives, characterized by mutual accusation (Bar-Tal, et al, 2014). In this line of thought, politicians are known to supply hateful narratives to their audience (Glaeser, 2002) influencing whom we hate and how much we hate to vary according to the political milieu

(Bar-Tal et al, 2014). It is not uncommon that hate speech is part of political discourse, not only among populist and extremist groups but also across the political spectrum and that information technology contributes to spreading and amplifying it (Council of Europe, 2021). In addition, hate speech is more likely to occur in the presence of group conflict and competition over resources, elections and regime change, high level of poverty and unemployment, and existence of a weak legal system (Benesch, 2014). Studies in Nigeria, for example, show that hate speech has been elevated to the status of a political campaign strategy (Ezeibe, 2021; Ezeibe & Ikeanyibe, 2017). During such contexts, people will be more vulnerable to fabricated news and conflict supportive narratives. Cognizant of this, the purveyors exploits individuals' propensity to join the disinformation and hate circle, playing as amplifiers and multipliers (Tandoc et al., 2017).

2.4.3 Consequences of Hate Speech

Hate is especially significant at the intergroup level, where it turns already devalued groups into victims of hate. When shared among group members, hate can spread fast in conflict zones where people are exposed to hate-based violence, which further feeds their hate. Hate can be reassuring and self-protective, because its message is simple and helps confirming people's belief in a just world (Fischer et al, 2018).

Through the construction of a common enemy myth, and portrayal of others in devalued manner, hate speech separates and excludes the targets from the community of humankind, reduces their social standing and acceptance, prevents them from civic and democratic participation (Simpson, 2013;Waldron, 2012), and increases their vulnerability to various forms of violence such as deportation, segregation, discrimination, and genocide (Benesch, 2014). Hate speech is thought to desensitize haters to the suffering of their targets (Soral et al., 2018) and hence contributes to genocidal acts, such as those committed in Rwanda, Bosnia, Myanmar, and Germany during the Nazi regime (Morrock, 2010; Smith, 2011). Targeting a community or individual because of their immutable or prominent characteristics slowly eradicates feelings of safety and security (Levin & MacDevitt, 1993). Considering the harm it caused to Rwanda and other countries, we can say that hate speech is the rising cause of bigotry and bloodshed. Hate speech coupled with disinformation regularly, if not inevitably, precedes and accompanies ethnic conflicts, and particularly genocidal violence (Benesch, 2014; Soral et al. 2018; Morrock, 2010; Smith, 2011; Timmermann, 2008). Understanding the link between hate speech and actual violence requires the understanding of mechanism that facilitates the violence. Most often, media and politicians are at the forefront of hate speech circulation. For example, RTLMC, also known as 'Radio Machete,' has

become the textbook example to elegantly describe how hate speech via media can cause actual violence (Richards et al, 2019; Benesch, 2008). Most often haters engage in name calling in a way that help them justify their hatred and possible violence against their target. Below, Table 1 shows some of the major genocidal incidents and the names victims were given before the actual violence:

Table 4 Name Calling in Genocide

S/N	Labeler	Targets	Nick name	Source
1.	Hutu of Rwanda	Tutsi of Rwanda	‘Inyenzi’/ cockroach	Morock, 2010
2.	Myanmar/ Burma	Rohingya	Kalar (Dark skinned) for Muslims and those supposed to be Indian descendants	Lee, 2021
			Bengalis (to mean they settlers who came from Bangladish)	
			Dogs, maggot, pigs, rapist	Morock, 2010
3.	Khmer Rouge regime leaders in Cambodia	Those opposing the regime	Microbes and a sickness (rot us from within” and hence must be removed)	Hinton, (2005) Romero, et.al, 2011
4.	Nazi, Germany	Jews	Rats ; hunted animals	Westermann, 2021

When we see the lives lost associated with hate speeches, between 1,500,000 and 3,000,000 Cambodians between 1975 and 1979, near to a million Tutsis of Rwanda between 7 April 1994 and 15 July 1994, nearly 5 million Jews by Nazi Germany, and 8000 Bosniak Muslim men and boys in July 1995 were killed by their labelers and their supporters. A mass genocide was imposed on the Tutsis by the Hutus. The genocide created an initial displacement of 1.7 million Hutus fearing reprisals, left 400,000 widows, 500,000 orphans, and 130,000 imprisoned on suspicion of committing acts of genocide (Tiemessen, 2004).

In order to see the link between the two, the following two models help understand the path. The first is the pyramid of hate which shows how hate transcends from simple to complex consequence, as seen in Figure 2.

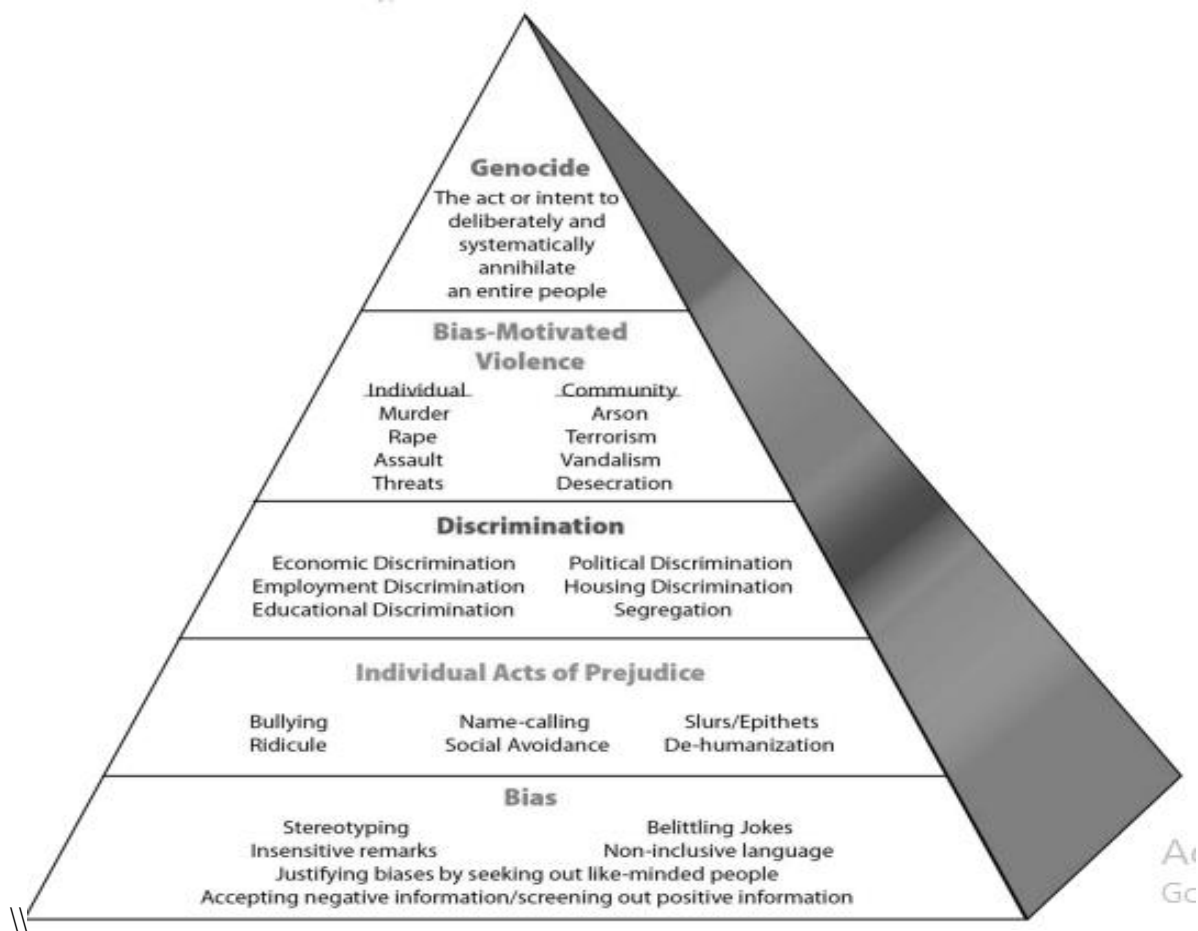


Figure 2 Pyramid of Hate

Source: Anti-Defamation League, 2005

The pyramid of hate tells us that as we move up, the severity and criminality due to hatred increases; while the most bottom one is more of daily ‘livable stereotypes.’ As we move up, we encounter prejudice at the bottom, hate in the middle, and finally violence including genocide at the top of the pyramid. The upper levels are supported by the lower levels as well as the majority of the problems are at the bottom which are more likely to be tolerated. If people or institutions tolerate or accept behaviors at the lower levels, it results in the behaviors at the next level becoming more accepted. Underneath a particular identity-based attack, we can see attitudes and inactions that set the stage for the final tragic outcome. From legal point of view, legal prosecutions are mandatory for the top-level hatred; civil prosecution for the middle ones; and public diversity education is prescribed as better response for the bottom.

A second model worth understanding is the theory of prejudice by Allport (1954). Gordon Allport identified 5 stages of prejudice (1) Antilocutin, (2) Avoidance, (3) Discrimination, (4)

Physical attack, and (5) Extermination. In this categorization, antilocution means “speaking against” which includes not only making jokes about another group (which is also called derogatory speech), but also the expression of hateful opinions (which is also called hate speech) (Allport, 1956).

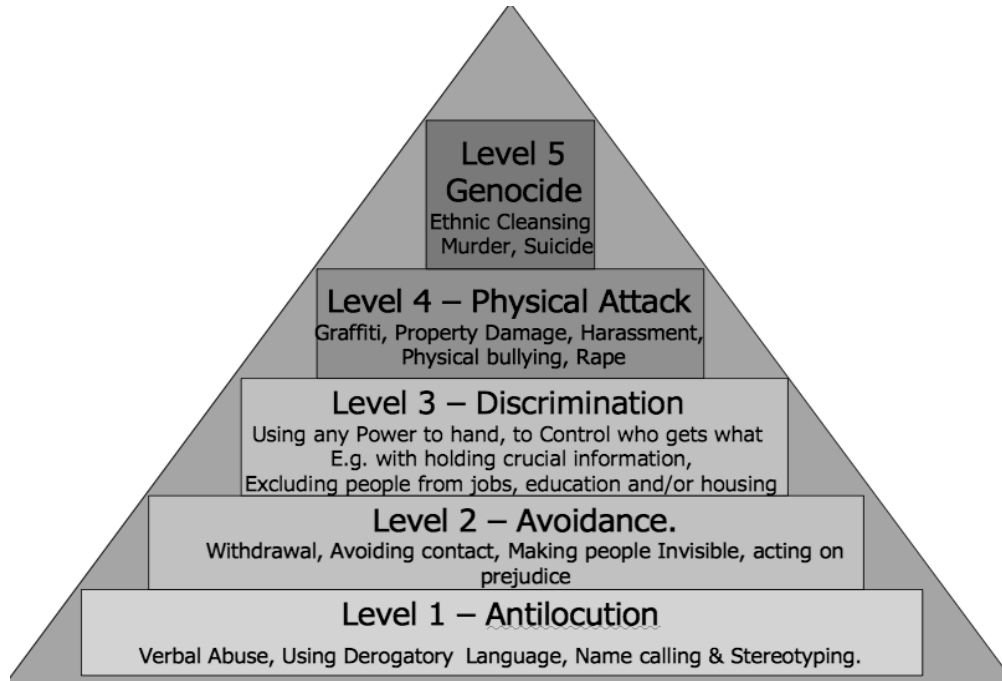


Figure 3 Allport's Levels of Prejudice

As people move on, nonetheless, such verbal attacks will lead to the next harsher physical consequences. The danger of hate speech emanates from the fact that it psychologically prepares members of the labelers to take some action against the labeled group. To illustrate this evaluative bias of language, Shiraev said to describe is to prescribe (2013). For example, calling a group as snakes implies the obligation to kill poisonous animal including the offspring. Because of its hostile nature, hate speech creates unnecessary factions within a society. Hate speech is a communicative phenomenon that not only disrupts intercultural dialogue but also contributes to the breakdown of multicultural societies (Wasilewski, 2019). The absence of hate speech can stop genocide (Skelly, 2022).

Nearly a decade ago, nonetheless, Gagliardone and colleagues contended that social media conversations in Ethiopia seemed to favor positive engagement over more aggressive or hateful forms of communication. What they meant, in essence was, these conversations did not reflect the ethnic and political polarization in the country as much as they anticipated. They hoped to mitigate some of the existing tensions by creating spaces of

engagement where ideologically opposed participants could communicate peacefully (2015, 2016).

Considering both the implicit and explicit attitude as well as the link between attitude and behavior, I proposed the following model of antecedent and consequence of hate speech.

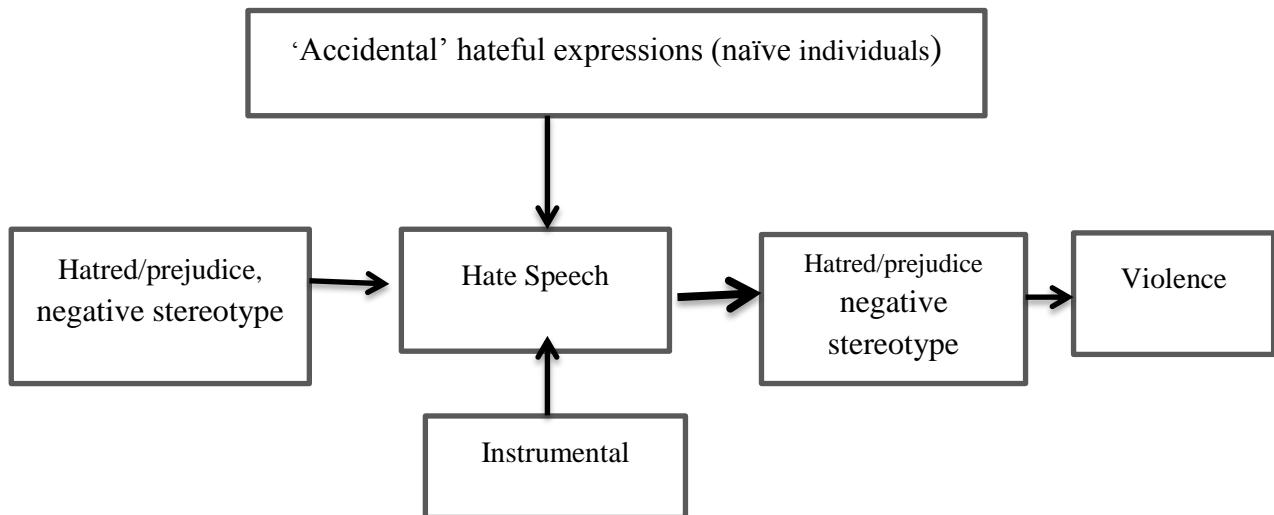


Figure 4 Proposed Antecedent-Consequence of Hate

Figure 4 shows hate speeches are more likely to be surfaced in three ways or reasons. The first one is due to hatred or prejudice. In this sense, hate mongers could be driven by the negative attitude they hold against their target. The second way is due to accidentally whereby naïve individuals can be drawn into hate speech, without apparent negative attitude towards outgroups. The third one is instrumentally wherein individuals could engage in hate speech due to some other reasons that they want to achieve via the hate speech. The reasons for the instrumentalist could be political or financial (hate speech for attracting supporters or for sale respectively). While the proposed model in figure 4 seems linear, violence has the possibilities of causing further hatred and negative attitude towards others. For example, studies show it is likely that perpetrators develop a stronger negative attitude towards their victims, who in turn have reasons to hate their attackers. Therefore, the model should be seen in cyclic way.

2.4.4 Actors in Hate Speech and their Techniques

Borrowing the idea of media ecology from Marshall McLuhan, Strate contends that the media hosts various stakeholders and constantly shapes cultures and perception, values, attitudes, and behavior of the social media users (2017, 2008). In times of political and social upheaval,

when insecurity prevails, ethnic leaders take advantage of uncertainty to consolidate their power, mainly for their own personal benefit (Carment, 2007). While competing over the minds of social media users, influencers employ a wide range of techniques, such as those known in persuasion, compliance, obedience, propaganda and related mind control techniques (Cialdini & Goldstein, 2004; Myers, 2010).

In order to convince their followers, elites use various techniques which mainly exploit the basic tendency of human beings (Cialdini, 2009), such as consistency seeking, self-esteem maximizing, group membership, terror manager, cognitive miser, foolish mistake makers, information seeker etc. (Baumeister & Finkel, 2010). First, they employ conformity techniques, which prompt individuals to change their stand due to real or imagined group pressure but not direct request, either to accord with others (Normative social influence) or to validate their stand (informational social influence) (Asch, 1955, in Myers, 2010). Influencers may also make direct request so that their followers would accept them, such as through flattering their target or playing scarcity and urgency techniques to convince audiences (Cialdini, 2009; Dolinski, 2016). Persuasive influencers are also experts in how to develop and employ such techniques as credibility, similarity and attractiveness, as well as the use of logic or emotional appeals (Caldini, 2009; Snijders & Helms, 2014).

Drawing lessons from political propaganda, elites could also employ techniques known in military indoctrination and related radicalization (Bradshaw & Howard, 2019; Silverstein, 1987). For example, the assignment of demeaning labels or offensive and pejorative names such as fascist, junta, or anti-democracy, separatists will be the strategy used to induce rejection or hatred in the minds of the audience (name calling). If repeated, name calling often leads targets to take on the names as reality and labelers to be desensitized for their victims. In what is named bandwagon, influencers appear as if they and their ideas are accepted by the majority and hence everyone else should join the group (Woolley & Howard, 2018). The bandwagon effect is part of a larger group of cognitive biases or errors in thinking such as group think (Forsyth, 2010). As a result, the bandwagon effect influences how posts are shared as well as interactions within online groups.

In its dangerous and abusive form, social influence can take the form of what is named as brainwashing/mind control, which is a coercive persuasion, mind abuse, thought control, or thought reform process (Tylor, 2004). In one of the most extreme forms, a phenomenon called gaslighting individuals are made to excessively rely on their source of information. Individuals could also follow a path that traps them to a fixed perspective that leaves them

with a tunnel vision. For example, they may adhere to a particular source of information as trusted and dependable one, rejecting every other source. Targets also prefer whom to be friend with usually leaning towards like-minded others. Deception and conspiracy theories are one of the most extreme but persistent examples of disinformation (Mahl & Schäfer, 2023).

Contrary to the fact, yet, many people believe that they are less likely to be under the influence of social media as compared to others, a phenomenon Pronin and colleagues called Introspection Illusion- our belief that social influence plays a smaller role in shaping our own actions than it does in shaping the actions of others (2001). As a result of this, conformity occurs nonconsciously and escapes our introspection.

2.5 Social Media: *(mis)use and hate speech*

Social media refers to a wide range of web-based applications that enable users to create, share, and exchange content while interacting with others, including but not limited to social networking sites (e.g., Facebook), blogs and microblogs (e.g., Twitter), internet forums, bookmarking sites, and online community sites (Kaplan & Haenlein, 2010). According to a recent study, today's social media users spend an average of two hours and twenty-four minutes per day across an average of eight social networks and messaging apps (Chaffey, 2022). The rise and popularity of social media is partly attributed to its ease of accessibility (cheap and easy), speed, interactivity, longevity, editability, scalability, trans-boundary (Fischer & Reuber, 2011). Preference to social media over traditional media is also attributed to the benefits it provides to users, such as safety needs, associative needs, esteem needs, identity or self-exploration and expression, and impression management (Krämer & Winter, 2008), voyeurism (exploring others' selves), offering and obtaining help, entertainment (leisure), and escape from stress. Social media is also preferred because it drives social change by empowering individuals and groups, giving voice to the voiceless (Ortiz et al., 2019). Social media also enriches diversity of opinion and allows citizens to access information that would otherwise be unavailable. Yet it also serves as sharing divisive and toxic rhetoric, disinformation, hate speech and polarization (Gallacher et al., 2021; UN, 2015; UNESCO, 2023), with a spillover effect on offline violence and protests and revolutions (e.g., Arab Spring, see Gallacher et al., 2021). In a Pew Research Center study, for example, 60% of internet users had witnessed offensive name calling, 53% had seen efforts to purposefully embarrass someone, and 24% witnessed sustained harassment of an individual (2014). In addition to providing an additional communication channel, therefore, social media

spaces increasingly shape today's conflicts and guide how they are handled due to their mobilizing, value-setting, and perception-shaping abilities (Mercy Corps, 2021). As a result, social media, once hailed as a force for freedom and democracy by facilitating debates and information flows and altering the nature of political engagement, has come under increasing scrutiny for its role in amplifying disinformation, inciting violence and extremism, and lowering trust in media and democratic institutions (Flore, et.al., 2019; Zeitzoff, 2017).

The internet allows for anonymous and pseudonymous communication, which can lead to people saying things they would never say in public (Citron & Norton, 2011; Gorka, 2019). Anonymity simulates deindividuation effects, which are defined as a loss of one's sense of identity and personal responsibility (Joinson, 2001). In addition, hating uses “snowball effect” which implies responding aggression with aggression, exactly what haters most often want (Gorka, 2019).

The second factor related to the internet is that, due to its low cost and potential for immediate revival, hate speech and fake news can remain online for an extended period. Hate speech online can be itinerant (Gorka, 2019). This also implies that, even if content is removed because it involves multiple actors, it may still be expressed elsewhere. The internet also enables non-journalists to reach a large audience. The rise of citizen journalism called into question the relationship between news and journalists, as non-journalists began to participate in journalistic activities and produce journalistic outputs, including news (Robinson & DeShano 2011, in Tandoc et al., 2017).

Online communication is typically asynchronous, allowing users to edit and reflect on their messages before sending them. As a result, we can assume that senders have greater agency and intent in the content they share online, which can help them present themselves in the best possible light (Walther, 2022).

When other technologies are added, social media platforms support manipulative disinformation campaigns carried out directly by individuals or by software known as bots, which are programmed to perform simple, repetitive, robotic tasks (Woolley & Howard, 2018). Bradshaw and Howard, (2019) found evidence of organized social media manipulation campaigns in 70 countries, up from 48 countries in 2018 and 28 countries in 2017. In their report ‘the global disinformation order’ Bradshaw & Howard (2019) indicated that Ethiopia is among the various countries taking part in social media manipulation.

Social media is also designed to provide users with search results that tend to confirm their opinions and views rather than results that represent a realistic distribution of opinions, which Pariser refers to as the filter bubble (2011). The filter bubble surrounds users with familiar ideas, leading to overconfidence in their mental frameworks. Because of filter bubble, an internet user only sees information and opinions that support and reinforce their own beliefs, as a result of algorithms that personalize an individual's online experience. Furthermore, it eliminates the opportunity to gain insight and knowledge that will change the way they think about the world (Pariser, 2011). The term "echo chamber" in news media refers to a situation where beliefs are reinforced through communication and repetition within a closed system.

Social media also emphasizes the bandwagon heuristic, as each post is accompanied by popularity ratings (Sundar, 2008). When a post receives a large number of likes, shares, or comments, it is more likely to be noticed by others and thus liked, shared, or commented on further (Lee & Thorson, 2008; Voorveld, 2019). As a result, social media popularity creates a self-fulfilling cycle that encourages the spread of false information.

Cyberspace is known for its hostile narratives, targeting feelings and emotions, and touching upon particular social vulnerabilities. Hostile narratives are designed primarily to go viral in such a way that topics are framed to reinforce community pride, instill fear, build a sense of urgency, or increase ingroup cohesiveness and outgroup hostility. In so doing, the narratives created are not always objectively false, nor are most of them classifiable as hate speech, but they are intended to reinforce tribalism, to polarize and divide, specifically designed to exploit social fractures, creating a distorted perception of reality by eroding trust in media, institutions, and eventually, democracy itself (Flore et al., 2019). Hate speech narratives frequently seduce the most in the "attention economy," fostering a sense of community and increasing the reach of such messages.

The last reason comes from the beneficiaries/audience perspective. In today's world, the younger generation, which has had extensive exposure to digital media, uses digital tools without considering what they are or how to use them (Hargittai, 2010). Social media users rarely verify the information they share, especially if the message comes from a socially proximate source—a set of layers with varying degrees of proximity to the reader (Kang et al. 2011). This is because socially proximate sources can help legitimize the accuracy of information shared on social media. This takes us to the point of referent power (French & Raven, 1959, in Myers, 2010).

2.6 Social Media Theories

Social media platforms like Meta, Twitter, and YouTube offer users a wide range of opportunities to both consume and create media content, signaling a potential shift in control over information, experiences, and resources (Barberá, 2020; Bossetta, 2018; Kaplan & Haenlei, 2010). Recognizing this, social media companies and social influencers struggle to win the attention and support of their audiences. In support of this, the Attention Economy Theory posits that people's attention is a valuable commodity in social media usage, leading platforms to compete for it. At the center of this competition, we find a large number of social media users whose media engagement varies in intensity, intention, goal, and outcome. This has attracted the academia in various fields like behavioral sciences, advertising, conflict studies, health, and security to explore why individuals use social media, their roles on these platforms, and how media engagement impacts them. As the result, a few salient theories are either developed or identified from other fields to shed light on user engagement, roles, and the effects of social media. While some of the theories are developed primarily to explain users' behavior, others are borrowed from related social science fields. Below I have categorized the theories into three based on the in which they are developed: media and communication, social psychology, and personality fields:

2.6.1 Theories from Media and Communication Fields

In the field of media and communication, five theories appear to be relevant in understanding the mutual influence between users and social media platforms. First, Cultivation Theory posits that much of our knowledge or beliefs stem from stories encountered in the media rather than personal experiences (Gerbner, 1998). This theory suggests that individuals form their attitudes, social norms, stereotypes, and values through repeated exposure to media content. Consequently, people may perceive the real world in a distorted manner, viewing reality through a 'television perspective'. When the internet is a hot spot for hate speech, this theory expect that more social media users who encounter such toxic contents are more likely to develop hatred. This is worrisome as, according to a study by the Pew Research Center in United States of America, a significant percentage of adult internet users (41%) have witnessed or personally experienced online harassment (2021). The theory adds heavy media consumers tend to respond rely on a cognitive shortcut in their thinking processes and potentially less reflective answers. This theory appears to be congruent with the 'mere exposure' theory of attitude formation known in social psychological science. In support of this view, Sternberg (2020) suggested that observers of hate content may transition from

passive spectators to active participants including trolling over time. Nonetheless, this theory seems it oversimplifies the intricate relationship between media exposure and audience behavior, overlooking the active role of individuals in interpreting messages. A related theory to cultivation theory is Agenda-Setting Theory, which refers to how the media's news coverage determines which issues becomes the focus of public attention (Groshek & Groshek, 2013). The central idea of agenda-setting theory is that elements emphasized by the mass media are more likely to be regarded as important by the public, the more prominently they are stored in people's memories, and that the more power they have on peoples' social and political vieews (Parmelee, 2013). Provided that political elites have the power to control which narrative to share dominantly on media platforms, this theory suggests they could control peoples' mind.

The third theory found in media and communication field is Use and Gratification Theory (UGT) by Katz and colleagues (1974). UGT holds that individuals actively, consciously, and purposefully seek out specific media platform and content to satisfy personal integrative needs (enhance status), affective needs, cognitive needs, tension release needs (escape from stress), and social integrative needs such as to interact with family and friends (Katz et al. 1974). In support of this, Ryan and colleagues (2014), used this theory as benchmark and identified relationship maintenance, entertainment, companionship, surveillance gratifications (e.g. glancing photographs), and escape (from worries) as reasons for using social media. This theory seems appealing as it assumes users have a reason for using media platforms (i.e., they have the agency) and hence helps to sort out why users behave the way they do on social media.

Spiral of Silence Theory, the fourth, is a theory in political science and mass communication that suggests an individual's perception of public opinion distribution affects their inclination to share their own opinions (Ho, 2016; Noelle-Neumann, 1993). It suggests that individuals are inclined to withhold their opinions if they believe their views are not widely shared by their peers or society (Ho, 2016). It is likely that people often look to cues from media and their social circles to assess the prevailing public opinion and adjust their behavior accordingly. When their opinions align with perceived majority viewpoints, they are more likely to speak out. Conversely, if they feel their views are in the minority, they may choose to remain silent. This dynamic perpetuates a cycle where majority opinions gain prominence while minority perspectives fade away (Ho, 2016). I argue that, this theory fails to explain two things: first, it fails to describe the difference between minority opinion and majority

opinion. In intergroup context, few in-group members could be enough to resist the views of a number of outgroups members. Second, this theory may not be a strong appeal to explain how few individuals stand alone with their views that they share online.

Fifth theory, ‘an Approval-seeking’ theory of online hate, suggests that individuals are mostly motivated to express hate online because they enjoy the interaction with like-minded individuals and seek social approval and attention from both known and unknown peers (Brady et al., 2021). It explains that individuals can form online friendships through posting hateful content and engage in hate activities to gain social gratification from online hate mongers who share their prejudice. In essence, this theory assumes that while hate messages may seem targeted at specific individuals, the main audience for such content is typically friends who share the same feelings of hate. This theory also assumes that for any hate monger who shares content, there are collaborators who enjoy the hate message.

2.6.2 Social Psychological Theories of Media Use

As social psychological theories primarily explain individual behavior in the context of others, some are powerful to understand the social media behavior of users. Social Identity Theory, for example, suggests people utilize social media to shape and affirm their identities and to establish connections with others (Turner & Oakes, 1986). In addition, this theory also explains individuals’ behavior on social media can aim at defending their collective identity or enhancing their collective self-esteem. SIT highlights that individuals often strive to present themselves in a manner consistent with their desired social identity and seek validation from their online communities (Luhtanen & Crocker, 1991). In support of this, Shared reality theory suggests that individuals are motivated by broad epistemic and relational concerns to achieve and maintain shared understandings of various objects in their social world with other people (Hardin & Higgins, 1996; Festinger, 1957). As William James suggested, notions of truth depend entirely on social validation: “Beliefs verified concretely by somebody are the posts of the whole superstructure” (James, 1907, p. 208). In other words, to be perceived as true, a belief must be shared with others. As social beings, we intuit the types of posts that will conform best to the prevailing attitudes of our social circle, such as indicated by the number of likes and shares by friends and family members (Council of Europe, 2017). This is congruent with Goffman’s Theory of Self-Presentation which maintains that most people strive to present themselves in a positive manner. For example, we tend to like or share things on social media that our friends or followers would expect us

to like or share (Goffman, 1959). There are evidences suggesting that affiliative motives may influence ideological beliefs to align with the progressive or conservative views shared within a given relationship or group, which may lead disproportionately to the adoption of system-justifying worldviews (Jost et. al., 2007).

Second, Social Capital Theory suggests that individuals on social media establish relationships and networks that can be advantageous to them (Johnston et al., 2013; Steinfield et al., 2008). This concept is rooted in the belief that people can leverage their networks to access resources, information, and opportunities that may otherwise be out of reach. Engaging with others' posts helps cultivate relationships, while liking and commenting on content serve as signals of respect and interest, contributing to relationship enhancement with friends and acquaintances (Ellison et al., 2014). This is perhaps well echoed by Social Exchange Theory which assumes individuals weigh the potential benefits of an interaction with another person against the effort and risks involved, in order to decide whether to engage or not. The theory suggests that individuals engage in social media interactions when they perceive the benefits to outweigh the costs (Surma, 2016). On social media platforms, exchanges can be based on likes, shares, comments, and follows. The fourth social psychological theory that explains social media behavior is Cognitive Dissonance Theory. According to this theory, when faced with conflicting beliefs, attitudes, thoughts or behaviors, people tend to reduce the dissonance created by bringing them into alignment with each other (Chakraborty et al., 2021). This theory argues that social media platforms provide an outlet for people to share their opinions and ideas on a large scale. As people are exposed to different perspectives, they may choose to modify their own beliefs in order to bring them into line with those of the majority (2019).

2.6.3 Personality and Social Media Use

While various factors influence an individual's engagement with social media, personality traits play a significant role in shaping their usage patterns. There are a number of studies with conclusion about the role of personality on social media usage, including disruption, frequency, and motives (Özgüven & Mucan, 2013). Extensive research has focused on the impact of the big five personality traits and the dark Tetrad on the online behaviors of social media users. Drawing on a very large community sample (N = 3835), for example, Sindermann and colleagues (2017) investigated personality differences between users and non-users of Facebook and found that Facebook users reported higher levels of extraversion

and lower levels of conscientiousness compared to non-users. In addition, narcissism has been associated with behaviors such as trolling and cyberbullying (Zannettou et al., 2019). Such links have significant implication for the circulation of hate speech.

2.7 Typologies of Social Media Users

Scholars in behaviors science and communication field have shown interest in exploring types of users on the bases of a wide range of schemes including time spent, modal behavior, motivation and needs, and patterns of engagement(O'Donovan, et a.,, 2013). For example, while Ostapenko and Sosniuk (2019) identified eight typologies (i.e., content generator, discussion initiator, active participants, spreader of content, imitator, conformist, observer, and inactive user), Cicek and Erem-Erdomuş (2013) identified five (e.g., inactives, sporadics, entertainment users, debaters, and advanced users), and Kim (2018) identified four (e.g., impression management type, lurker type (passive observers who leave little trace of their presence), enjoyer and relationship focus, and social value orientation type). Mirbabaie and Zapatka (2017) identified information starters, amplifiers and transmitters as the main roles played by social media users in the context of crises communication.

With regard to the time spent on social media, “active users” are those who engage actively in social media platforms through regular posting, commenting, sharing, and often initiate conversations and contribute actively to discussions. In the context of hate speech, evidence show more visible Twitter users (with more followers) are more likely to become targets of hate (ElSherief, et al, 2018). “Trolling” is another typology of users characterized by the tendency to harass, threaten, embarrass, or derogate other online or offline. With regard to those users who offend others trolling, spamming, cyberbullying, and hate mongers are used in varying contexts such as disinformation (Zannettou et al., 2019), provocation and harassment (Mkono, 2015), and harassing and insulting (Cheng, et al., 2017). In addition, various types of trolls are identified in previous studies. For example, while Shachaf and Hara conducted interviews of Wikipedia trolls, finding themes of boredom, attention seeking, revenge, pleasure, and a desire to cause damage to targets (2010), Narchuk identified two types of trolling, single and collective trolling which implies trolling as an individual and trolling in a group respectively (2020).

With regard to Ethiopian online hate speech, Gagliardone and colleagues (2016) classified the communication in the context of hate speech online as “going against” and “going toward”. The former consisted of statements that represented conflict-maintaining behavior or advocated hatred, incitement or discrimination, while the later consisted of statements that

helped to maintain a communicative relationship by acknowledging the other person's or group's position and by creating engagement with other members in the conversation even if the tone was critical.

Hateful users tend to be more densely connected and their content spreads faster farther, and to a wider audience compared to non-hateful users (Mathew, 2018). As the haters' aggressive actions continue and develop, the haters are increasingly likely to devalue their victims (consistent with self-perception theory). That is, they trust their own simplistic and flawed thinking, blame their problems on others, and project their own deficient character onto others. 2018). hateful users are densely connected, and are "power users" in the sense that they tweet more, in shorter intervals, 71 times more likely to retweet others hateful posts, and 41% focus to send the tweet to another hateful users (p-values < 0.01) (Ribeiro et al., 2018).

Research on users' online behavior appears to vary in terms of methods (such as single platform versus multiple platforms) and contexts (such as political, educational, advertising, etc.), with many studies focusing on motivations, frequency, and common activities. However, there appears to be a lack of empirical studies on how social media users in Ethiopia specifically respond to online hate speech, particularly in a context where identity politics plays a significant role in the propagation of hate speech. It is more likely that how individuals react to hate speech could potentially impact whether such messages contribute to peace or conflict. Therefore, addressing this gap could help make informed interventions to counteract hate speech in Ethiopia.

2.8 Countering Hate speech

Considering the pressing consequences of hate speech globally, nations, media companies, and other organizations working on democracy and freedom of expression have called for measures to be taken to counteract before it is too late (Justitia, 2020). Nonetheless, there is a lack of consensus among the stakeholders on what kind of measure to take against hate speech as well as variations in the level of commitment demonstrated to fight it back. Below a review of actions taken by different stakeholders are presented

2.8.1 Legal Means

Various governments are increasingly defining hate speech in their criminal codes in an attempt to directly regulate dangerous speech, hate speech, and/or harmful rhetoric both on and offline (Justitia, 2020). In East Africa alone, for example, countries such as Kenya, Rwanda, and Ethiopia are known for having hate speech laws, though with a varying degree

of its application. The Ethiopian parliament, for example, adopted a proclamation in March 2020, requiring social media companies to remove hate speech and disinformation within 24 hours of receiving notification. The proclamation, nonetheless, does not detail how this expedited removal process will be operationalized. The law also states the consequences of disseminating hate speech which are found illegal. Article 6 of the Proclamation also outlined that sharing hate speech would not be deemed illegal if it falls under academic research or scientific exploration, political commentary, artistic creation, performance, and religious teaching (FDRE, 2020). Nonetheless, it has received fierce criticism from a number of stakeholders and scholars. Ayalew for example argued: ‘The proclamation-exhibited ambiguity in the definitional part and contain provisions lacking adequate clarity and specificity as well as overly broadened stipulations’ (2022, P 6).

Even though countries like Ethiopia has criminalized hate speech, criminalization of hate speech is confronted with criticism on moral, legal, and practical grounds. The debate at global level, which took over six decades, is full of several principled arguments, both for and against hate speech laws. Below the main views of proponents of free speech and those that promote hate speech laws are presented.

Proponents of free speech take moral, legal, and intellectual appeals and argue that individuals shouldn’t be prevented from expressing their idea and feeling even hatred (Schauer 2015). First, they assume freedom of expression is a fundamental human right that even helps the protection of other rights, as articulated in Universal Declaration of Human Rights (UDHR) and Article 19 of the International Covenant on Civil and Political Rights (ICCPR). From moral aspect and intellectual appeal, unlimited freedom of expression is importance for the discovery of truth (Sunstein, 1993; Post, 2011) and also is a sign of respecting the right of the speaker as well as right, agency, ability, and autonomy of the listeners to choose what to listen and believe (Baker, 1997). More importantly, while such laws can be used by politicians in power to suppress their opponents (Wibke, 2008), they may also impose a significant enforcement burden and not necessarily reduce hate speech in all contexts (Gelber & McNamara, 2015)

Proponents of hate speech legislations on the other hand ask ‘what if others are exposed to danger because of the unrestricted speeches?’ Through the pretext of freedom of expressions, evidences show, individuals and groups can be exposed to various forms of violence such as exile, isolation, discrimination and even genocide (Waldron, 2012). In a democratic society,

they hold that, freedom is subject to formalities, circumstances, restrictions, and penalties, according to the law (Council of Europe, 2017). This takes us to the hate speech law (Amdur, 1980; Brison 1998; Greenawalt, 1989). International criminal law and Article 20 paragraph 2 of the International Covenant on Civil and Political Rights (ICCPR) requires States to prohibit certain severe forms of “hate speech,” including through criminal, civil, and administrative measures (Article 19, 2015). This is when it rises to the level of (1) advocacy of national, racial, or religious hatred that (2) constitutes incitement to (3) discrimination, hostility, or violence (Article 20). In addition, the UN’s Rabat Plan of Action outlines a six-part threshold test to guide such determinations, namely: (1) the social and political context (2) status of the speaker (3) intent to incite the audience against a target group (4) content and form of the speech (5) extent of its dissemination and (6) likelihood of harm, including imminence. This shows the need to put a limit to free speech.

The first step in any argument for criminalizing hate speech is, therefore, to establish that such speech is outside the protective ambit of the right to freedom of expression (Howard 2019). Hate speech laws are a complex and contentious issue, with a key debate centering on the balance between freedom of expression and the need to protect individuals from discrimination and persecution (George, 2014). Most scholars and practitioners argue that effective action against the use of hate speech requires: recognition of the fundamental importance of freedom of expression, tolerance and respect for equal dignity; identification of the conditions conducive to the use of hate speech and taking appropriate measures to remove them; and the involvement and commitment of a wide range of private and non-governmental actors, in addition to public ones (European Council, 2017).

Nonetheless, the recommendation by United Nation Strategy and Plan of Action on Hate Speech seems appealing (NU, 2020). This recommendation divides hate speeches in three levels of severity and provides different recommendations: a must legal prohibition, a probability legal prohibition, and protecting the hate speech (no legal restriction; lawful hate speech). Lawful ‘hate speech’ for example, are recommended to ‘be protected from restriction under Article 19(2) of the ICCPR, but nevertheless raises concerns in terms of intolerance and discrimination, and merits a critical response by the State’ (Article-19, 2015, P. 18).

2.8.2 Educational Means

The educational means, which focus on prevention aspect, capitalizes on the need to empower citizens to constructively use media technologies including detect and constructively respond to hate speech and disinformation (UNESCO, 2023). Several experts observed that by enhancing media and information literacy (MIL), individuals can learn to effectively navigate social media content, recognize the influence of media in their lives, appreciate different viewpoints and goals, and consequently develop critical thinking skills when engaging with media platforms (Article-19, 2015; Council of Europe, 2023). Media and information Literacy (MIL), nonetheless, cannot be the silver bullet to eradicate hate speech but that a wide range of societal responses are required (UNESCO, 2023). Still, educating and promoting the practice of critical reflection and verifying facts, while also taking measures to eliminate purveyors of hate speech, are additional strategies to combat the issues of hate speech and misinformation (Pennycook & Rand, 2019). In educational means to fight back hate speech could be benefited from involving elites. Siegel and Badaan, for example, observed that elite-endorsed messages that prime common identity and tolerance helps in reducing the spread of hate speech online (2020).

Teaching the values of peace and diversity as well as how to disagree and debate in a manner that is reasonable and yet civil is also an important way to address some level of hate speeches. In addition, it is recommended that social media users be instructed that freedom to post comments and thoughts online is not a license for profanity, harsh comments, or hate speech (council of Europe, 2021). Under certain contexts, it is also shown that users can be benefited from engaging in counterspeech, which is practice of responding to speech that seems harmful or offensive. Challenging hate speech, debunking or critiquing harmful speech, amplifying alternative viewpoints, providing accurate information, and fostering empathy and understanding are also found to be constructive ways of countering hate speech (Jääskeläinen, 2020).

2.8.3 Working with Social Media Platforms

Legal measures are also applicable with regard to monitoring media intermediaries such as Meta, Twitter, and YouTube which have rules (community standards) against hate speech. Even if such social media platforms pledge to remove hate speech contents, the reality on the ground are far below than the expectation (Carlson & Rousselle, 2020). This failure has implication for the rise of hate speech as its consequences. During the conflict in the North, for example, Facebook was found that it failed to remove hate speech which are supposed to

attract genocide level violence (Crises Group, n.d). It was also caught falsely reporting to have taken down harmful contents while they are still visible on the platform (Bureau of Investigative Journalism, 2022; The Observer, 2022). Yet, closely working with such platforms and demanding them to abide by the laws can help remove harmful contents including hate speech (Carlson & Rousselle, 2020).

It has to be noted that platforms' action against hate speech content is dependent on the participation of users in reporting hate contents to it. With regard to Meta's action from July-September 2021, for example, a report shows users' participation in reporting hate speech to the company is significantly lower (3.5%), as compared to the identification and action made by its own workers which accounts for 96.5% (Meta, 2021). Empowering citizen to participate in reporting hate speech to media companies and responsible national entities are helpful in countering the problem.

Chapter Three

Methods

The aim of this dissertation is to investigate the various types of ethnic-based hate speech expressions in Ethiopia and the mechanisms involved in their occurrence. In addition, it also aims to explore the reactions of social media users to hate speech posts so as to infer how their roles contribute to the rise of hate speech. Social media contents are the data analyzed to address the research questions, supported by memos and interview transcripts. This chapter, therefore, presents how the study was carried out so that the findings and conclusions can be understood in light of the approach used. It presents the design employed along with its justification, data source and sampling techniques employed, data collection processes and techniques used, dataset formation and clearance, data analysis techniques, trustworthiness issues, ethical considerations, and finally bias reduction mechanisms employed.

It has to be noted that the order of the subsections presented below doesn't necessary imply the actual steps followed in the research process. As a grounded theory study, many of the research process are not linear rather iterative, going back and forth as situations demand. The topics presented below are ordered, therefore, merely to conform to the 'standards' of writing up a research report.

3.12 **Study Design:** *Grounded Theory Study*

Grounded Theory is 'a set of procedures for the systematic analysis of unstructured qualitative data so as to derive by induction a supposition that explains the observed phenomena (APA, 2015, P. 471). Since its inception in 1967 by Straus and Glaser, what was initially proposed as a single approach has undergone numerous epistemologically positioned revisions forcing the originators into accepting it as 'a cluster of suggested techniques but not commandments' (Straus & Corbin, 1998, P. 4). Other GT scholars also accepted it as a family of methods, approaches, or methodologies that are and should be flexible, idiosyncratic, and contextualized (Guetterman et al., 2017; Babchuk, 2011; Glaser, 2003, 1978). Part of the manifestation of its flexibility and remodeling is that it can be used as purely quantitative design (Glaser, 2008), a purely qualitative design, quan-qual design (i.e., MM-GT, Glaser & Strauss, 1967; Guetterman, et al., 2017), as well as qual-qual approach (Halaweh et al., 2008; Charmaz & Mitchell, 2001).

Whatever revisions and remodeling are called for, nonetheless, inductive coding, memoing, adherence to the constant comparative method, simultaneous data collection and analysis, the

use of theoretical sampling, theoretical saturation, and ongoing theory development remain the hallmarks of the design (Charmaz, 2006; Urquhart, 2013). The analysis in GT involves concept labeling, categorizing, finding relations among categories, and generating a midrange theory from such relationships (Charmaz, 2006; Schreier, 2012). In so doing, this non-linear and self-correcting approach enables the identification of the main attributes of a phenomenon within its setting and the core social-psychological process involved (Charmaz, 2006; Urquhart, 2013).

This dissertation employed GT design because the study of hate speech lacks independent theoretical foundation as well as it aims to generate conceptual categories that can both describe and explain how hate speech manifests and occurs on social media. More specifically, this design is suited to generate concepts and mechanisms that can both describe and explain the types and nature of hate speech and the mechanisms involved in its occurrence. More specifically, the study employed the constructivist grounded theory (CGT) of social media content analysis which is already suggested by Charmaz for social media analysis (2006). CGT is preferred as human behavior and action (i.e., hate speech in this context) is based upon the meaning that individuals place on people, things, and situations and how such meaning is interpreted and communicated through language (Charmaz, 2017). In addition to helping develop explanatory concepts and models, CGT also allows for the generalizability of the findings (Charmaz & Thornberg, 2020).

Charmaz proposed a constructivist approach to Grounded Theory's content analysis which is now widely used while conducting content analysis when using the Grounded Theory methodology (2006:10). She contends that 'Archival records and written narratives, video and photographic images, internet posts, and graphics may give you insights into perspectives, practices, and events not easily obtained through other qualitative methods (2006, pp. 39).

It has to be noted that this study started as a Qualitative Content Analysis (QCA), until some limitation observed during pilot test prompted changing it to CGT. The early findings observed in the pilot study (using QCA), for example, provides surface level descriptive categories of themes that lack details on explanatory concepts, and forced interpretation to be limited to using the actual text as evidence (Sandelowski, 2000). For example, while QCA forces inferring codes and themes only based on textual or pictorial evidence, memoing showed some roles could be inferred not from actual texts but from the overall scenario

analysis (keen field observation). These gaps call for a different method than QCA, and CGT is a possible remedy for these gaps.

Nonetheless, signs of QCA are still traceable in this work, especially for the research question on the roles of media users. Indeed, as a descriptive method, QCA is suited as a supplementary rather than a standalone technique (Bryman, 2012). In support of this, Suddaby in the article 'From the Editors: What Grounded Theory is Not', elegantly stated the following about the relationship between CA and GT:

... both techniques can form part of grounded theory studies. The important difference is that grounded theory describes an overall method for systematically gathering and analyzing data, but content analysis describes a specific context within which a distinct type of data can be gathered and analyzed (2006, P. 5).

In addition, I believe existence of methodological integrity between the two. Principal areas of epistemological and methodological compatibility between the two include both draw on a constructivist epistemology that fits well with an interpretive approach (Schreier, 2012; Annells, 2006), their suitability (and integration) to social media content analysis (Cho and Lee, 2014; Lai & To, 2015; Halaweh, 2018; Pousti et al., 2013); and being based on naturalistic inquiry that entails identifying themes and patterns, involving rigorous coding (Cho and Lee, 2014); the aim of developing analytic density, employing an iterative approach to both data collection and analysis, and using constant comparative method (Priest et al., 2002). Even if making such a shift can cause some deviations from the standardized protocols, the shift is an indicator of being attentive to idiosyncratic use of designs- a criterion for rigor in assessing methodological trustworthiness and a sign of methodological self-consciousness (Levitt et al., 2017; Urquhart et al., 2010).

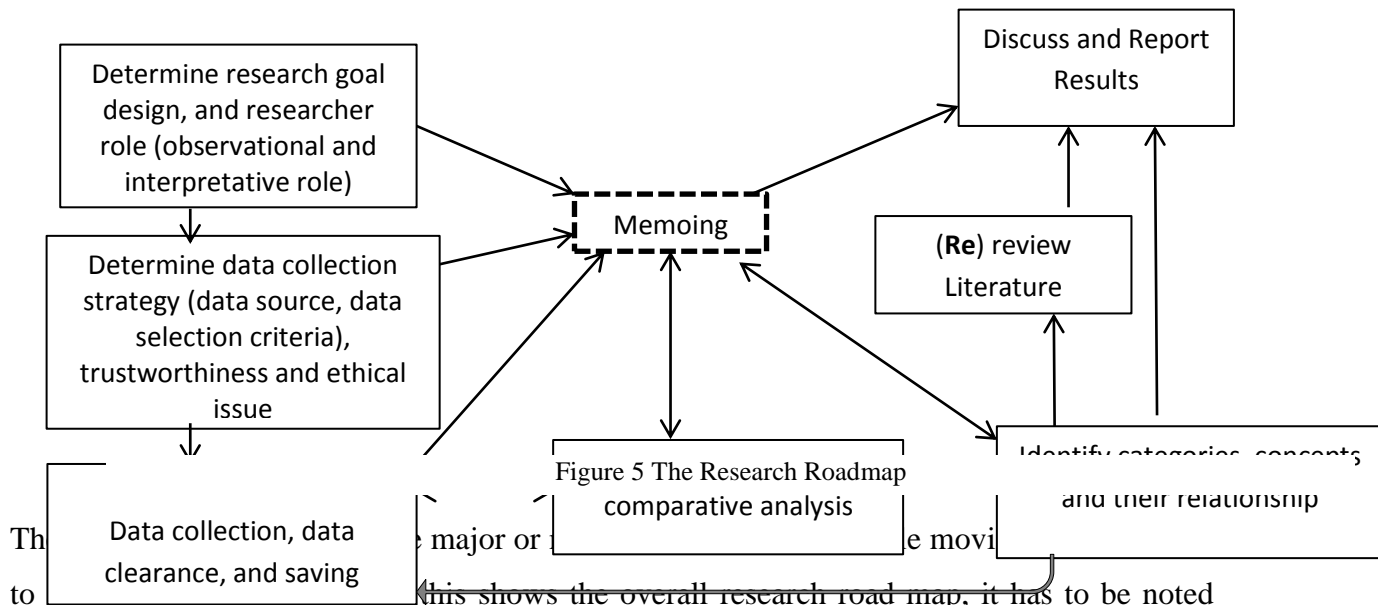
3.2 Philosophical Assumptions

Epistemologically, the study closely attends to social media contents by extending the 'field visit' for more periods (July 2021 to January 2023) so as to enhance sufficient data collection. Because the researcher can only understand hate speech within a broader framework that is contextually positioned within a certain time and situation, ontologically this study assumes a relativist position and hence the existence of multiple realities (Charmaz, 2006). The paper reports these multiple realities using the actual words of the data sources. As shared reality is interpreted by the researcher and that '...reality arises from the interactive process and its

temporal, cultural, and structural contexts' (Charmaz 2000, p. 523), methodologically, the study takes an interpretive stance. More importantly, the study used inductive logic, studying the topic within its context. With regard to axiological assumptions, the study admits being value-laden nature of the work, such as recognition of subjective values deployed in the overall research process (Glaser & Laudel, 2013).

3.3 The Research Road Map

The overall research step followed in this study is presented in figure 5.



this shows the overall research road map, it has to be noted that there are reversed steps, such as going back to the previous steps in almost each of the parts. For, example, changing and modifying a research question made after reaching at the analysis of data. Needless to say, GT is self-correcting process open to change and modification to be made when needed. At the center of the figure is memoing, showing it is related to virtually every process and decisions made on the research. While some literature was reviewed before collecting data, a significant portion of it is added after concepts and categories are emerged. The re-reviewed or added literature part aims to make sense of the generated categories, in the discussion part.

3.4 Sources of Data

While the majority of the analysis is done on social media contents, additional data are collected from key informants. The data sources are presented below:

3.4.1 Social Media

Due to the ability to efficiently gather and analyze large amounts of naturally occurring hate speech data (e.g., text, images, and video) with their context in an unobtrusive and replicable

manner, this dissertation relied heavily on social media contents (Halaweh, 2018; Kim, 2018). As grounded theory researchers seek to avoid contrived settings, publicly available social media content help in obtaining reach contexts in which the content are framed in. Due to the grounded theorist's focus on the explication of social processes (Glaser & Strauss, 1967), an understanding of the conditions that bear upon participants is critical. For this, natural settings, such as social media, allow for the documentation of contextual influences on social actions or behavior (Lindlof & Taylor, 2011). In the same vein, as people have a pervasive motive to conceal their prejudice and negative intergroup attitudes (Sommers & Norton, 2006), it makes studying hate speech using self-report methods less reliable (Fisher et. al., 2018). As such, relying on social media, where people feel free to express bigotry, is a more suitable way of studying hateful expressions (Matamoros-Fernández & Farkas, 2021).

In addition, social media contents provide the views and online behavior of the elite and ordinary citizens from various camps, which in turn helps get the big picture of the phenomenon that can portray the issue at the national level. To this end, a significant part of the data for this study is collected from three social media platforms: Facebook, YouTube news channels, and Twitter which are the first, second, and fifth most used social media platforms in Ethiopia in 2022 (Statcounter.com, 2022). According to this report, Facebook accounts for 74%, YouTube for 13.5%, and Twitter for 3% of social media usage in Ethiopia. Even though Twitter is the fifth most popular social media platform, the authors had the impression that the platform is mainly used by high-profile individuals.

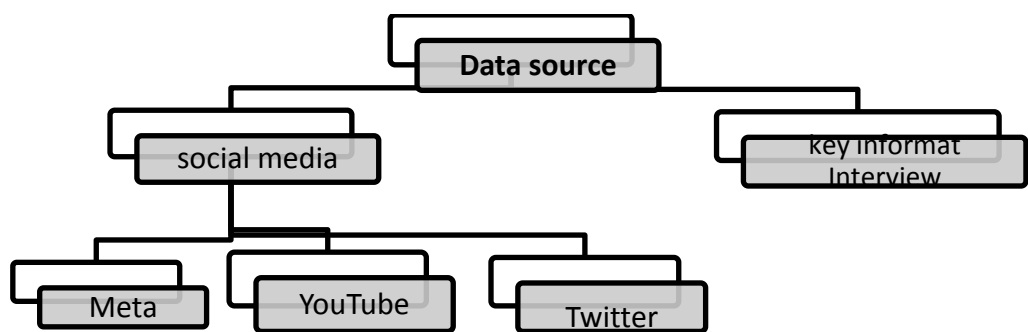


Figure 6 Data Sources

Identifying these three platforms is not sufficient, though. To collect hate speech content from these platforms with relative ease, identification of and reliance on hot spots is very crucial. Hot spots, in this study, are public spaces of social media accounts or pages that the researchers identified as possessing a lot of hate speech content during the data collection

period. To this end, ten Facebook accounts, four Twitter pages, and four YouTube news channels are purposefully identified and followed for a certain period. Table 2 shows the 18 account holders who are followed as hot spots:

Table 2 Hot Spots

Platform	Account holder (code)	TA.D	TA.B	EL.B.	ABR.D	TFI	ZE.B	AC.T	GT	DW	AS
Facebook	≠ followers	370k	170K	104 k	156 k	109k	124k	126k	126k	151k	162k
YouTube	Account holder	Tigray Media House		Abbay Media		Kello Media		Reyot Media		-	-
	≠ subscribers	144k		381k		112k		131k		-	-
Twitter	Account holder	H.A		K. G		Gi.Gu		Dan		-	-
	≠ of followers	118k		120 k		95K		114K		-	-

Three reasons are identified for selecting these account holders. The first is that the page or account should be created and used for three or more years before the data collection to obtain rich hate speech contents. The second criterion used to select these accounts and pages is the popularity of the users as measured by the number of people following them (more than 100,000 for Facebook, 50,000 for Twitter, and 100,000 subscribers for YouTube). Having a huge number of followers is taken as a criterion, as, through tagging, sharing, and retweeting, their public space is more likely to contain rich hate speech data. The third criterion used is using at least Amharic and English as their means of communication. Yet, the actual contents collected for the analysis are those framed only in Amharic and English.

While relying on the above-mentioned criteria to select hotspots, care was also taken so that the selected accounts or pages (implicitly the specific contents collected) portray ‘the big picture regarding ethnic-based hate speech in Ethiopia, rather than reaping skewed data from a particular ethnic group that could negatively affect the conclusion the paper would make. To this end, accounts or pages from the following four camps are included: those believed by the researcher to have an ‘extreme’ form of nationalism, and those who appear to be voices for their respective ethnic groups, preferably ethnic Amhara, ethnic Tigray, and ethnic Oromo. I chose individuals from these ethnic groups (who use Amharic and English in their social media communication) as emerging local studies show members from these ethnic groups are heavily involved in hate speech (Skjerdal & Mulatu, 2021; Ayele, 2020).

At this juncture, nonetheless, it is necessary to bear in mind that the owners of accounts or pages identified as hot spots are not necessarily the producers of the hate speech content found in their public space. They might post their content, save or share others' hate speeches (including those that target their group), be tagged, or host discussions in which guests offend others (for YouTube news channels). In addition, they are also identified as hot spots because their posts and shares (including positively framed ones) are usually seen to attract hate-fueled discourse and/or hate speeches.

Hence, the hate speech content collected and analyzed in this study were produced by either the account holders themselves or others, including high-profile politicians, activists, scholars, religious leaders, ordinary social media users, and Ethiopians in the diaspora. Hence, individuals with no social media account at all are also part of the producers of the hate speech analyzed in this study. This is because hate speeches they produced offline can be recorded and aired online. It has to be noted that, social media platforms are also rich with content which were already disseminated in other media platforms like mass media and print media.

3.4.2 Key Informant Interviewees (KII)

Six KIIs are conducted with theoretically selected informants who are believed to have experience in closely observing, documenting, and/or fighting back against hate speech in Ethiopia. For this purpose, interview guides with open-ended items are prepared in a semi-structured format that is believed to provide dependable and rich qualitative data through probing based on participant responses (Galletta, 2013; Adams, 2015). The interview guide included an introduction part that describes the purpose of the interview and an oral informed consent part. Had the method allowed face-to-face interviews with producers of hate speech, they would have been contacted to validate the findings. Yet, this study relied on social media archives, and hence the actual producers of them can not be easily reached at. This is because of both the inaccessibility of the individuals as well as their unmanageable size. Yet, the actual social media users whose contents were analyzed could not be found to validate the findings. As a remedy for this gap, therefore, the following individuals who are actively working in the area of hate speech are contacted.

- i. **Ethiopian Media Authority (EMA)** is an autonomous government organization, with a mission of Promoting Pluralistic and Responsible Media through enabling regulation, capacity building, and partnership. Four informants from this office working at different department are interviewed.

- ii. **The Centre for the Advancement of Rights and Democracy (CARD):** is a not for profit organization registered in Ethiopia under the Civil Societies Law, and has been working to promote and defend human rights and create democratic governance in Ethiopia. From this organization, two staffs relevant to this research are interviewed.

Table 3 Interviewees' Profile

S/N	Age	Sex	Edu.lev	Organ.	Position
1.	48	F	LLB	EMA	Vice director for legal affairs
2.	33	M	MA	EMA	Internet and social media monitoring and capacity building v/director
3.	30	F	MA	EMA	Communication Affairs
4.	50	M	BA	EMA	Research and Dissemination Office Expert
5.	35	M	MA	CARD	Advocacy and Communications Manager,
6.	28	M	MA	CARD	Social Media Monitoring

The purpose of the KII is double. The first is to collect fresh, relevant data to substantiate the findings. Second, which is the primary purpose of the interview is to validate the findings obtained. The primary aim of the interview is hence to check the findings obtained from social media content and determine whether emerging concepts and categories resonate, and are recognized by them as part of 'member-checking' process (Scott, 2011). As the interview progressed, for example, the questions became more focused on collecting specific data needed to 'confirm' the categories or concepts generated earlier, and enrich the categories, thereby assisting the development of conceptual model (Charmaz, 2006).

3.5 Data Collection Procedures

3.5.1 Judging Contents as Hate Speech: to identify and collect ethnic-based hate speech content (written words, pictures, or video format) from the hot spots, a broad working definition for hate speech is used (see definition of terms on page 13). In addition, a checklist in 'Yes' or 'No' format was also prepared (based on the working definition) and used to decide whether particular social media content can be judged as hate speech or not (see Appendix-6).

Nonetheless, the screening process is open to any possible hate speech content which may not fulfill any of this point. This openness is opted as hate speech expressions can be disguised within historical discussions, humors, or any emerging styles.

The study has also included fake news, provided that it has a hate speech essence. Nonetheless, the decision has been made by attending both to the manifest as well as the latent content meaning of communications. For this, context and scenario analysis beyond the

actual text are taken into consideration. More importantly, the decision is made by putting oneself in the shoes of both the object and the subject of the hate speech. Nevertheless, because of the interpretative stance and the belief in multiple realities, some of the contents judged ‘hate speech’ in this study may not be considered as such by others.

3.5.2 Inclusion Criteria: in this study, only hate speech that is ethnically-based, framed in Amharic or English, posted in the last three years during data collection, and found only in the public space of the hot spots (Facebook, YouTube, or Twitter) is collected and analyzed. Hate Speech that target other protected characteristics than ethnicity is not included.

3.5.3 Time Frame for Data Collection: Because data collection was responsive to emerging codes, concepts and categories, the research timeline needed being flexible (Glaser & Strauss, 1967). Overall, the data collection took 17 months, from July 2021 to January 2023, but with some periodic breaks for data analysis, writing and reflection on the work done. Initially, a flexible schedule was followed in such a way that each identified hot spot was browsed once a fortnight, from July 2021 to February 2022 (pilot). As time went on, the data collection time frame became more and more flexible. Nonetheless, even if ‘the actual time frame’ for data collection ended in January 2023 because of the felt saturation, contents considered ‘new arrivals’ are also entertained, provided that the data found has an add-on element.

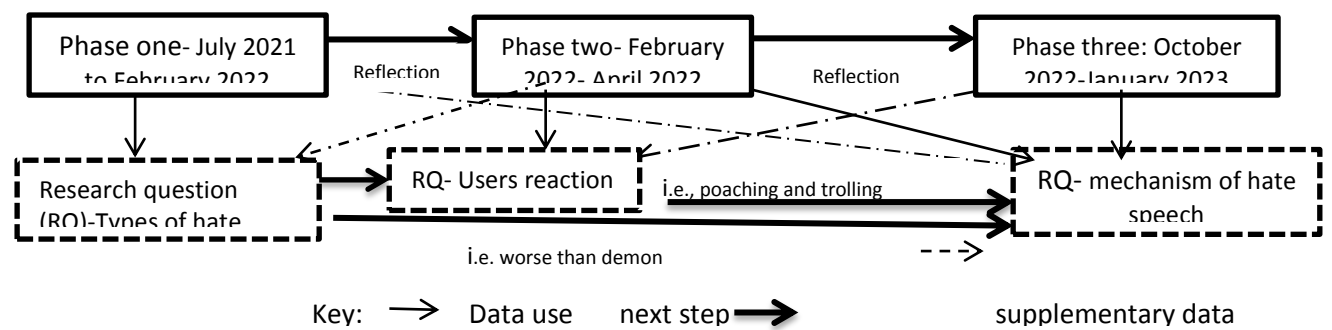


Figure 7 Data Collection Period and Major Tasks

From July 2021 to February 2022, data was collected and possible codes and categories pertaining to types of hate speech were generated. Yet, this was at a very initial stage (piloting), and hence there was a need to add some more data both to refine the categories and enrich them with more attributes. The data collected from February 2022 to April 2022 were meant to refine the categories and concepts generated in the first step (for types of hate speech) as well as to address the research question regarding users’ reaction to hate speech posts. The data collected from October 2022 to January 2023 was meant to address the last

research question- mechanisms involved in the occurrence of hate speech. As shown in the diagram above, the categories and concepts on the mechanism of hate speech are partially informed by the concepts generated by types of hate speech (e.g., worse than demon) and users' reactions (e.g., trolling). As a result, previously generated concepts and categories are refined, expanded, and applied to the next research questions. As the pilot test was presented in October 2022, the reflection made on the pilot has helped to seek more data, especially for the third research question. With regard to such work, specific focus to an objective of the research at a time and after finishing it giving the attention to the next objective is recognized and recommended by Charmaz who said,

‘Quite often, we discover that our work suggests pursuing more than one analytic direction. Thus, we may focus on certain ideas first and finish one paper or project about them but later return to our data and unfinished analysis in another area’ (2006, P 10).

3.5.4 Data Collection Techniques

Technique of collecting the needed data from the hot spots is a key issue worth reporting. In this study, both manual searching and automatic scraping techniques are applied to collect relevant social media contents from the identified hot spots.

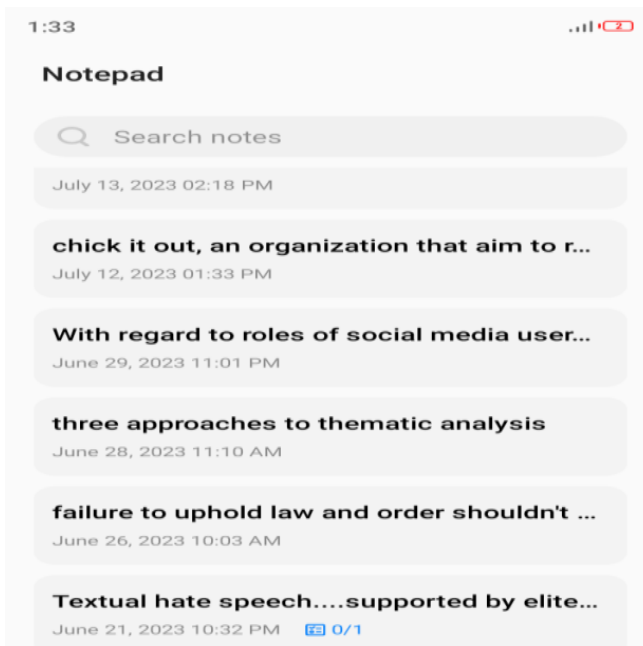
3.5.4.1 Manual Searching: This study heavily relied on manual searching of the contents of the identified accounts and pages. Since abusive behavior, such as toxic expressions, is relatively infrequent, random sampling results in datasets are more likely to be extremely skewed towards benign samples (Founta et al., 2018). Hence, by scrolling down accounts or pages and examining their public archives, it was possible to come across hate speech or hate-filled content in various formats, such as texts, videos, pictures, or photos. While texts and pictures are copied and pasted; the videos are transcribed and formed dataset. When the selected hate speech post has more than 100 comments, the first 100 are manually examined (the remaining are not!). If the comment given to a hate speech post is less than 100, all are examined, and those found to be hate speech are copied and pasted on the data set. The researcher opted to limit the search for the first 100 comment as, otherwise, the bulky nature of comments may distract data management.

Because manual searching is a tiresome endeavor, the following greasing mechanisms are used to facilitate the process of finding the needed hate speech data beyond reliance on the hot spots: The first is liking or subscribing to hot spots so that once they are liked or

subscribed, the content they share appears on the account of the author and hence facilitates the finding of more relevant data. The second grease used is using keywords and hashtags for Twitter accounts. Hashtags (#) using keywords and phrases known to devalue the target groups are used to facilitate the identification process of hate speech; they become linked to all of the other tweets that include that word or phrase, along with the contexts (Twitter.com). Hate speech videos were downloaded using Facebook video downloader, YouTube video downloader, and Twitter video downloaders, and then saved and transcribed for further scrutiny. Social media has lots of texts without contexts that may mislead one to interpret them outside of their context, and clues as from where and when the data come from, who participated in shaping them, and the true intention of the sources of the data may or may not be found. Therefore, the data collection aimed to capture a range of contexts, perspectives, and timeframes (Charmaz, 2006).

Beyond manually collecting the social media content, the approach also helped the researcher to be in the 'field' for an extended period (i.e., the observational model of social media). This observation allowed the researcher to see how hate speech is constructed and constrained by contexts (Starks & Trinidad, 2007) and to narrow focus on aspects of a setting or on specific activities and interactions.

It has to be noted that data collection is not limited to the planned time schedule mentioned earlier. Rather, when the researcher encounter, relevant information while not 'on desk', the data is saved on the researcher's Facebook storage feature (including on the cellular phone) so that it could be examined later on. The 'notepad' in the researcher's cellular phone (TECNO Spark 4) also helped in memoing any relevant popping up ideas that came to his mind. The following screenshots shows these techniques:



3.5.4.2 Automatic Searching: In order to examine the role of social media users in responding to hate speech posts, any comment (positive or negative) have to be collected. To this end, 14 ethnic-based hate speech posts each with more than 1,000 comments by the time they were visited were collected from the public space of four purposefully selected YouTube news channels and four Facebook accounts. Then using a web assistant called ‘export comment.com’, 100 comments from the main post is extracted automatically. This technique- ‘export comment.com’- automatically extracts 100 random comments (for free and more comments with payment) given to the hateful post, along with the profile name of the person who commented, profile ID, date on which the comment is given, the number of likes, replies, and reactions each comment received, and the site/link it could be retrieved with. In order to address the research question, therefore, 1400 comments from 14 posts were collected, step by step. The majority of the main hate speech posts from which the comments are extracted are videos, while the remaining are long hate speech texts. After extracting the comments, they were saved in excel format for further scrutiny.

It should be noted that the 14 hate speech posts from which comments are extracted are found from four Facebook and four of the YouTube accounts (eight of the 18 hot spots, no comment from Twitter platform is analyzed). As the researcher moved from collecting comments from one hate speech post to another, saturation dictated where to cease seeking for more comment, in this case 14 posts from which 940 were found relevant for the analysis. Even if the initial dataset of comments consisted of 1,400 random comments, 460 of them were considered irrelevant as they didn’t allow for inferences from them or help address the

research question. These removed comments include unclear symbols and marks (e.g.,!,?), incomplete sentences that do not give meaning, as well as words written in languages other than Amharic and English (e.g., Arabic words). Finally, personal identification, such as users' names were removed from the dataset. Hence, after removing such irrelevant information, the data analysis was made based on 940 comments.

3.6 Theoretical Saturation: is the phase of qualitative data analysis in which the researcher has continued sampling and analyzing data until no new data appears, and hence all concepts generated are fully developed and their linkages to other concepts are clearly described. Because the analysis in this study aimed at both enriching the attributes of a category or a concept generated as well as reaching an explanatory level, the data collection was kept until the researcher felt no new code, concept, or category could emerge. Indeed, this understanding is taken as the primary goal of researchers employing GT (Charmaz, 2006). For this, the data collection and analysis were made back and forth manner (Creswell & Poth, 2018). To find more relevant data and context, a continuous cyclic process of data collection, analysis, coding, memo writing, and direct theoretical sampling is employed. In this decision, the researcher examined multiple incidences and perspectives to identify and deal with possible gaps and variations (Glaser & Strauss, 1967).

3.7 Dataset Formation and Data Clearance: the collected data, including interview transcripts and memos, formed a larger corpus of data that helped organize and examine them with relative ease and present evidence for the claims made. In all these processes, data screening and checking are thoroughly done. For example, any irrelevant items and users' names and pictures are either removed or obscured. In addition to the memo developed, this dissertation used, 725 hate speech posts (a single word, shorter phrases, sentences, paragraphs, lengthy texts, and pictures), 24 hate speech video transcripts, 940 comments, and six interview transcripts.

3.8 Data Analysis

The data analysis employed suggested techniques known in GT, while at the same time being open to flexibility in response to the actual work. As a principal qualitative data analysis, inductive coding is employed (Braun & Clarke, 2006). Step by step coding processes (initial, axial, and theoretical coding) were passed along with, memoing, and comparative analysis. Because texts that tell the story behind other texts at least suggest the social context for the analysis (Chamaz, 2006), the interpretation thoughtfully considered the context in which it is framed, including the time. Because interpretivism entails subjectivity, the researcher put

himself in the shoes of both the subject and object of the hate speech analyzed so as to judge how the content under question can be considered as hate speech.

The data analysis is not a one shot action; rather takes iterative process that involved a cyclical process of refining or tweaking the latest version of coding and categorizing. The overall data analysis took multiple periods of observation, data collection, memoing, and analysis. As data analysis proceeded, therefore, it shaped the next data collection focus, allowing a check on the relevance and appropriateness of the interpretation and providing more and more description and explanatory power to the identified categories and concepts. In addition, at several points feedback was collected from the research adviser and colleagues (including presentation of earlier findings at research conferences). The researcher has ensured that the analysis is based on rich, empirical, original, and trustworthy data that reflect in-depth the psychosocial process, which are characteristics of a good GT (Aldiabat & Navenec, 2018). The main steps and techniques employed in the analysis are presented below:

3.8.1 *Memoing*: throughout the research process, the researcher recoded his progressive thoughts about the research questions, data issues, his initial impressions and interpretations, the emerging codes, and developing concepts and categories. While some of the reflective thoughts are used in making decisions on the research process, others are included either to substantiate the findings or to help readers make sense of the findings and conclusions made. unstructured sample memo are presented in appendix 5.

3.8.2 *Familiarization and Context identification*: observing the social media field is an integral part of GT studies when done on media contents (Charmaz, 2006), which helped the researcher to be familiarized with details of hateful contents on the social media, their contexts, and the activities and interactions that take place among users (Guba & Lincoln, 1982). As presented earlier, the overall data collection period was extended to be around 17 months. This period included the ‘field visit’ or observations done on social media, than merely the actual or formal data collection one. Therefore, both as a social media user and a researcher, the ken observation made for such longer period helped me to be rich in hate speech contents, to better comprehend social media behavior, which in turn reduced the ambiguity that might have been created during coding and analysis. In addition, I also took note (memo), just similar to ‘analytical field note’ which is common in ethnographic studies (Bryant & Charmaz, 2012). The dataset formed and the memos recorded hence helped the researcher to inspect them several times, ensuring adequate level of familiarization that assisted the analysis process.

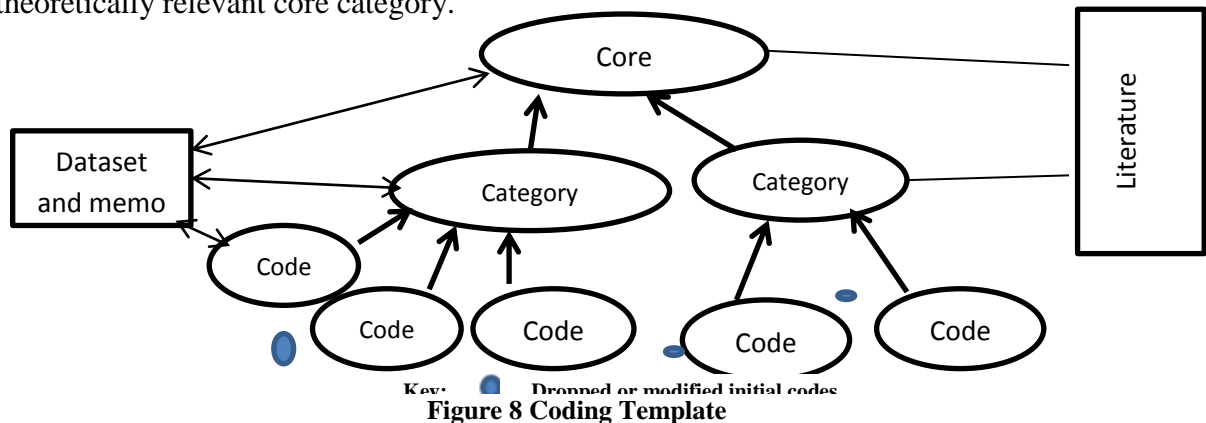
3.8.3 Coding: To his best knowledge, the researcher approached the data collection with an inquiring mind and persistence to find rich data that can give rise to detailed, focused, and complete codes. Taking interpretative stance, the data collection was also accompanied by interpreting the content right on the spot, and within the context it was embedded. Contexts prevented the researcher from being led astray- interpreting data outside of its context. While coding is done simultaneously with data collection, the dataset formed helped the researcher to immerse into it, and become familiar with each datum, and to go back and forth with reflective thoughts, so that new codes can be found along with editing and modifying codes. In addition, immersion and the ease of having the ‘cognitive map’ helped to sense what kind of additional data to seek.

While coding, the researcher began by being open to what is said, by whom, to whom, and under what context, and then pursue my hunches and potential analytic ideas about them. Explanatory notes (memos, examples, and descriptions) are also created to explain the codes emerged and concepts generated and the contexts in which they were framed. The coding process follows an iterative process where an initial stage of coding is followed by more abstract coding that elevates codes to super-ordinate categories and, finally, the construction of a core category with a robust explanatory power (Charmaz, 2006; Kenny & Fourie, 2015). More specifically, initial coding, axial coding, and theoretical coding (2014) are conducted as described below:

- ✓ ***Open /initial coding:*** is the line-by-line coding of each account in isolation (e.g., significant words or phrases), where I explored and examined fragments of data for their essence with possible theoretical category a specific datum suggests (i.e., theoretical sensitivity). The initial codes identified, nonetheless, were provisional and informed holes and gaps in the data at hand and the analytic process. For example, after transcribing a hate speech video, I reread it until I acquire a sense of the direction for analysis. Then I conducted line by line coding and labeled certain concepts common to them. While doing this, I also set aside those codes which have little relevance to the research questions as well as those that appeared infrequently.
- ✓ ***Axial coding:*** here I examined how a code is related with other codes (i.e., through constant comparative analysis) and created tentative categories, which are then constantly audited to confirm their superordinate status. By going back from categories to dataset, memo, and ‘field’ for more focused data, each axial code is explored in detail so that more and more attributes of each conceptual category that makes it unique is delineated

(Charmaz, 2006; Strauss & Corbin, 1998; Birks & Mills, 2012). As the constant comparison went on, it prompted further questions to be explored via repeated scrutiny. I then tested the categories for referential adequacy by returning to the raw data. The concepts and categories identified are then named, defined and thoroughly described with definitions and illustrations.

- ✓ **Theoretical coding:** here similar categories were merged to give rise to the final theoretically relevant core category.



To obtain accuracy of evidence in the category, establish the generality of a fact, and reach higher levels of abstraction and conceptualization, comparisons between empirical data and concept and categories, between concept and categories, and among data were conducted during data collection and analysis. With regard to coding for hate speech types, for example, the analytic work on the taxonomy described ended once I am convinced that a good fit is present between the data and the categories that make up the taxonomy. This implies the taxonomic elements account for all of the data analyzed. Finally, I used theoretical sampling to delineate the properties of a category; to check hunches about categories; to saturate the properties of a category; to distinguish between categories; and to clarify relationships between emerging categories.

As illustration, the coding process for roles of Social Media Users initially identified 36 codes, which were later reduced to 27 codes by deleting some and merging similar ones. Once the final codes are decided, similar codes are merged into six categories of roles. Yet, as analyzing the categories continued, a decision was made to delete two and split another into two themes. This gave rise to five final categories, typologies of users.

It has to be noted that, the data analysis didn't give equal or equivalent focus to each code or 'issue'. For example, some issues appear to be theoretically important in terms of providing

explanatory power more than others and hence took more time in the analysis. In addition, the analysis is not done in straightforward or linear manner. For instance, I opted to dwell on a few salient contents (e.g. tracked some trolls), for a certain period of time to see any developments relevant to the study, to identify as much attributes as possible.

In general, even though not in mutually exclusive way, the three research questions were addressed in somehow sequential manner- in order of the first, the second, and then third research question. That means, addressing the first research question helped having some clues to address the next research question. After identifying the four types hate speech, for example, I invested on ‘worse than demon expressions and found earned reputation ‘DNA’ that helped the focus of the analysis pertaining to the mechanisms of hate speech. By the same token, after identifying the five roles of social media users, I then focused more on trolling and finally found the ‘DNA’ for poachers and their techniques (paving ways for the analysis of the last research question).

In all the analysis, greater care was taken so as to make the analysis not unduly partial, while at the same time dwelling on salient codes, categories, and concepts that are theoretically relevant to the research. Outliers are well recognized only if they are theoretically significant instances.

After completing the analysis, I started comparing the concepts/categories I developed with the relevant literature. Indeed, reviewing the literature is used during data collection, data analysis, and discussion. The literature helped me develop sensitizing concepts, generate plausible explanation for the findings, clarify contributions, and indicate avenues for future research. This gave me the opportunity to show how my analysis fits, extends, or challenges leading ideas in the field. While describing the finding, quotes and excerpts are used to illustrate the characteristics of the themes under discussion, and that interpretation is part of the description of each theme and its illustrations.

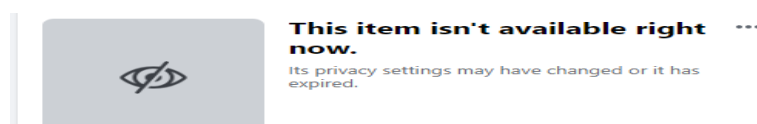
3.9 Trustworthiness

This dissertation has executed the necessary cautions needed to make the study process and its results as trustworthy, dependable, auditable, and transferable as much as possible so that the conclusions arrived would resonate to the ethos of the scientific community. To this end, the following measures are taken: First, prolonged engagement with the datasets as well as the ‘field’ was made to achieve credibility, which is the “fit” between views of text sources and my representation of them (Tobin & Begley, 2004; Lincoln & Guba, 1985). The

researcher, for example, made prolonged readings of the data collected along with the contexts in which they were observed to gain an impression of their content as a whole and to begin to generate ideas, hunches, categories, and themes that interpret the phenomena. Second, evidence of systematic methodological decision making provides a justification for the reader (Guetterman et al, 2017). As to my view, the analysis is executed in transparent, authentic and detailed manner. The researcher, for example, made coding and the analysis process clear with sufficient details so that others can understand how and why decisions were made (Braun & Clarke, 2006; Koch 1994; Loreli et al., 2017). The researcher also acknowledged the influence of his background, perspective, and assumptions including possible bias and remedies taken (see section 3.11).

Third is *confirmability* which requires the researcher's interpretations of findings derived from the data, requiring the researcher to demonstrate how conclusions and interpretations have been reached (Tobin & Begley, 2004). For this, the researcher worked in a systematic manner to form conclusions and interpretations to be grounded in the data (Stiles, 1993). For example, the researcher provided a thick and rich description of categories. The study also tried to adhere to the assumption that the broader social context influences how meaning and intentions are inferred from texts (Boyatzis, 1998).

Throughout the process, the researcher was open to any information coming from the data throughout the entire process and continually engaged in reflexivity (memoing), which is keeping a self-critical account of the research process, including the internal and external dialogue, to facilitate understanding of both the phenomenon under study and the research process itself (Charmaz, 2006; Watt, 2007). For a clear audit trail, the researcher has kept records of the raw data, the transcripts, and a reflexive journal or memo to systemize, relate, and cross-reference the data. For example, each code and category has a corresponding illustration or quote traceable back to the original data. All transcribed videos (except one that is removed by the source before being downloaded) are downloaded and are available for checkup. Nonetheless, while all of the hate speech contents analyzed are available in the dataset, a few of them may not be available online as they might be deleted, removed, expired, or restricted by the owner. For such items, the search using their web links reads the following:



Triangulation across different data sources and available theories reviewed is one of the basic strategies used to enhance quality and trustworthiness. The use of constant comparative analysis resulted in a type of triangulation of findings (Lincoln & Guba, 1982) and a way of delineating the conceptual boundaries of categories (Denzin, 1971).

3.10 Human Subject Protection

Even though it is crystal clear that ethics is and should be at the heart of all social science research, there is limited guidance on application of these ethical principles when using online social media research, which is wrought with ethical dilemmas (Fossheim & Ingierd, 2015; Markham & Buchanan, 2012).

The data collected in this study are from public space of the hot spots identified, without the need to directly communicate with human subjects on the social media. It is assumed that data existing in public, which did not require password to get access to the material, cannot be considered as human-subjects, and hence bring little ethical challenge with regard to informed consent (British Psychological Society, 2013; Pace & Livingston, 2005). Nonetheless, as part of respecting the privacy of the actual owners of the data and other relevant ethical issues, the following measures are taken. As reminder obtaining informed consent is neither practicable nor mandatory. As the majority of social media users (source of the contents used in the study who are in thousands) are either inactive or not easily reachable to communicate with, it doesn't require seeking informed consent. Nonetheless, the research applied anonymity to avoid any unintended harm, including judgments or ridicule, and to preserve or protect professional reputation (especially the high-profiles who can be takes as representatives of their ethnic groups).

Hence, personal identifications of sources of data (e.g., names, photos) were not made visible (are deleted or obscured) while presenting excerpts or quotes. In addition, while presenting the exact words of sources is necessary, the paper used the translated version of quotes or shortened it by changing some words so that they could not be retrieved in a way that could expose individuals to any unintended harm. Even if web links can be presented along with pictures and quotes, I assumed they could expose participants to unintended harm. This is mainly used when I feel that individuals (i.e., elites) are vulnerable. For some of the unidentified social media users, links are provided when I believe it helps the credibility of the finding. At some point, I also removed or obscured the ethnic names to avoid further misunderstandings and fueling conflict in the future. I have also obscured slurs that insult ethnic groups, such as by omitting or shading a letter from that word. Therefore, unless

necessary in special circumstances (if concerns or reasons emerge later on), the majority of this research work doesn't focus on who specifically said each word but rather on what is said on social media. In a nutshell, the basic ethical expectations such as anonymity, informed consent, the need to protect three key entities namely the community, the researcher, and sources of data from possible and unintended harm the privacy-public nature of data are well respected and applied with regard to the key informants.

3.11 Researcher's Background, Role, and Possible Bias

I am a teacher by profession and a psychologist by training. I am a member of the Ethiopian Orthodox Church. Born to Oromo parents, I grew up in Afar region where I had the opportunity to befriend with individuals from various ethnic and religious backgrounds. I am a father of two.

I feel I am non-partisan, with little interest and involvement in Ethiopia's politics (except for feeling worried for what I feel is a wrong sociopolitical situation). Well, this is what I consciously thought about me. Taking into consideration the fact that identity-based politics affects self-concept and identity formation, I feel that the currents of ethnic politics is less likely not to influence the thoughts and feelings individuals have about 'us' and 'they' and as well as what we are seeing on the social media and offline. Therefore, researching ethnic-based hate speech may have some impact on the way I see and make sense of the socio-political environment. With regard to contents, I judged as 'hate speech', for example, I understand that others may treat them as 'not hate speech'. This partly is due to the complexity of the meaning of hate speech as well as bias that creep into interpretation or perspective I and other observers take. So, I admit incursion of possible bias into the data collection, analysis, and conclusions made-knowingly or not. As possible indication for the incursion bias in data collection, analysis, presentation etc, the following excerpt is about a feedback from my advisor, given during the pilot study's report document), on Wed, Jun 22, 2022

I think you are a little biased about the Geda being insulted by extension an insult on Oromo, you need to withhold some of your feelings. I know all of us have some of it but as this is academic writing more balance is needed. I know in qualitative research, bias is there but yet balance is vital.

In addition to showing possible bias in my judgment, this kind of comments and many more that I received from my advisor have helped the paper to be vigilant for manageable biases.

I would also say that writing on hate speech, especially in the context of the current Ethiopia, is a daunting task, or at least not an easy task! In addition to affecting the findings, studies show repeated exposure to hate speech could lead to the development of negative sentiment towards others. As such, the attempt to be a 'hate speech expert' could lead one to be a hate monger. The remedy for this is perhaps entering the work only when one is sure that he or she has insulations such as deeply held values to human diversity and peace, as well as the ability to continually reflect on one's thoughts and feelings and make necessary measures to guard off possible incursion of negative sentiments.

As a researcher, nonetheless, my aspiration was to inspect what is said and how and why it is said with sensitivity, creativity and feeling that give qualitative research its soul. To this end, I tried my best to limit my role to observer and interpreter, while consciously monitoring and controlling biases. In order to avoid manageable and unacceptable bias from creeping into my work, therefore, I used the following checks and remedies indicated in Table 4:

Table 4 Safety Check to Guard off Bias

S/N	Expected sources of bias	Action taken
1.	<p>Screening contents: minimizing bias while deciding and screening processes. This include taking contents that are free speeches(e.g., Mere critics or expression of dissents), and unduly taking hate speech posted by other ethnic groups than those framed by members of my ethnic groups</p>	<ul style="list-style-type: none"> • Use definition and checklist • Take perspectives • Consider context • Academic integrity and honesty • Fair selection across groups- balance the ethnic composition of the hot spots (i.e., avoiding taking too much hate speeches instances framed by a single ethnic group that may unduly skew the findings)
2.	<p>Interpretation of contents: ensuring interpretation is based solely on data at hand and not on preexisting feeling and ideas. Question whether</p> <ul style="list-style-type: none"> • Interpretation is biased against ethnic-politics; or an ethnic group • There is differential interpretation of hate speech targeting my own ethnic group and other ethnic groups. • Am I too emotional when I came across contents against my ethnic group? • Do I justify or trivialize hate speech against other ethnic groups 	<ul style="list-style-type: none"> • Memoing and ongoing reflection; • ‘Members checking’, advisor and KIIs • Present evidence for each claim • Demonstrate academic honesty • Don’t justify but report • Make user the interpretation is based on the data and the context • Perspective taking, putting on the shoes of others
3.	<p>Presenting finding. Biases enter when using examples to illustrate ideas.</p> <ul style="list-style-type: none"> • Do I frequently present hate speeches that target an ethnic group; while portraying another ethnic group as haters? • Have I relied on rare instances as dominant theme (this may give a wrong impression of groups)? 	<ul style="list-style-type: none"> • Distribute examples of hate speech across ethnic groups (i.e., illustrations or examples of categories and concepts should evenly be distributed across groups, if possible). • Dealing with rare instances. Avoid presenting rare instances as theoretically or practically significant. Rely on referential adequacy

Chapter Four

Findings

This dissertation aims to explore the types of ethnic-based hate speech expressions and how they occur and propagate in Ethiopia. The research questions guiding this investigation were: a) what are the types of ethnic-based hate speeches circulated on social media (typology of hate speeches)? b) how do social media users react to ethnic-based hate speech posts? (typology of users with regard to responding to ethnic-based hate speech posts), and c) how do ethnic-based hate speeches occur on social media?(mechanisms involved in the occurrence of ethnic-based hate speeches). The answers to these questions are critical to comprehending the complex dynamics of how hate speech is framed and circulated on social media, as well as inferring the offline interethnic dynamics in Ethiopia. While a significant amount of data for this dissertation came from social media archives in text, picture, and video formats, six interviews were also conducted to substantiate the findings. In addition, memoing is meaningfully used to shape the data collection and analysis.

This chapter organizes the findings systematically into three overarching thematic parts based on the research questions posed. In Section 4.1, findings pertaining to the first research question are presented. This includes the four core categories—dehumanization, enemification, devaluation, and attacking and repelling—as well as their possible relationships. Each of the core categories and subcategories is presented with quotes that illustrate the embedded concept in each as well as the context in which they are found. Section 4.2 presents findings pertaining to the second research question. As such, trolling, pace-making, peace-making, informing, and guarding are presented, along with excerpts illustrating the concepts implied. Finally, Section 4.3 presents the findings for the last research question—mechanisms involved in the occurrence of hate speech—with evidence from the dataset that substantiates the claims made.

It has to be noted that the findings are presented in reversed steps as to the data analysis steps. That means that while codes and subcategories first emerge and core categories are formed later by combining similar subcategories, the finding section presents first a core category followed by its subcategories. This finding section presents only the findings generated from the data, so discussions, interpretations, and implications are presented in the next chapter.

Section 4.1 What Are the Types of Ethnic-based Hate Speeches Circulated on Social Media (typology of hate speeches)?

The analysis made of the data regarding this research question shows hate speech manifest itself in four types: dehumanization, enemification, devaluation, and attacking and repelling. These are presented below, along with their subthemes, contexts, and illustrations.

4.1.1 Dehumanization: *Animals, Objects, and Demons or even worse!*

The first hate speech core category identified is characterized by expressions that describe an ethnic group as other forms of being than a fully human being. This category includes hate speech expressions in which ethnic groups or their members are portrayed as similar to or worse and less than harmful or despised animals, worthless or harmful objects, and demon- which are named in this study Animalization, Objectification, and Demonization respectively. Each of these is presented below:

4.1.1.1 Animalization

The data analyzed shows some of the hate speech expressions that portray ethnic groups as pet animals, wild animals, poisonous snakes, rodents, and insects. While some expressions use generic names, such as animals or herds, others use specific types of animal species to describe their target.

Table 5 Animalization Expressions

Animalization Illustrations	Target group and Main Contexts
<p>Animals used in the expression: ‘daytime hyenas’, ‘Bugs’, ‘chameleon’, ‘Serpents’, ‘worse than beasts’</p> <p>Sample Excerpts</p>	<p>Target: Tigrayans and their subpart (e.g. TPLF)</p>
<p>Animals used in the expression: ‘Animals’, ‘herds’, ‘worse than beasts’, ‘lower than animals’,</p> <p>Sample Excerpts</p> <ul style="list-style-type: none"> ‘is worshiping trees practiced in this modern era? ■■■ is animal. 	<p>Target: Oromo and its subparts</p> <p>Context: ruin of Shashemene³ town, the (alleged) internally displaced people from Wollega, and Irrecha Celebration</p>
<p>Animals used in the expression: ‘Apes of Ras Dashin mountain’, ‘donkeys’, ‘dogs’, ‘herd’, ‘</p> <p>Sample Excerpts</p> <ul style="list-style-type: none"> ‘We don’t expect more from donkey’s descendants You your 	<p>Target: Amhara and its subpart</p> <p>Context: The conflict in the North.</p>

³ a town in Oromia Region which saw serious killing of civilians and property destruction the day after a prominent artist Hachaluu Hundessa (of Oromo descent) was killed on Monday June 30/2020.

Table 5 presents hateful expressions that use ‘animals’ and its derivatives, the ethnic group mainly referred to in the expression, and the contexts in which the expressions are framed. As the table shows, the common element that is threaded in such hateful expressions is using ‘animals’ to portray their target, and hence, the conceptual territories of such expressions is named ‘Animalization’.

Table 5 also shows that diversified animal species are used to refer to targets. The following two pieces of evidence show the animalization types of hate speech under dehumanization. In one of the videos analyzed, for example, an apparently Amhara man and presumably critic of the leading political force in the Amhara regional state labeled presumably all ethnic Amharas directly and blatantly ‘donkey’, *rationalizing* his labeling on the ground that they don’t understand things around them (Elias, 2020). The expression from this hate monger apparently implies donkeys do not understand things around them and so does the target of the hate speech. In the second excerpt, another social media user seemingly from ethnic Oromo shared a picture on his Facebook page satirically portraying Fano as donkey, as seen in post below:



(Source:<https://www.facebook.com/photo/?fbid=5235709903193774&set=a.339568829474597>)

The caption on the above post written in Amharic could be translated as ‘Eritrean government is creating job opportunities for Fano, and they are registering for the chance.’ The content was posted by a social media user who was later found to be a poacher against ethnic Amhara, who is apparently showing resentment against unemployed Fano’s involvement in

the conflict in the North or Fanos' presence at all. The satire the source used is that he mentioned 'Fano' in the expression but showed a donkey's picture—apparently to mean Fanos are donkeys. The difference between the two animalization expressions that used donkeys is that the first directly attacked ethnic Amharas indiscriminately, while the latter separately attacked Fano. In addition, the first expression provides implicit reason as to why he used donkey in the expression, while the second expression doesn't tell anything about why they are referred so. During the conflict in the North, Fanos were supporting government forces, yet only to enter into fighting them soon after the conflict in the north is over. While the fighters and their supporters insist on using this name, Fano, the Ethiopian federal government refers to them as 'rebel force' and 'Jawsa'.

Animalization is also well noted against ethnic Oromos and ethnic Tigrayans. In a video that recorded a meeting and went viral later on, for example, an apparently Tigrayan man was heard saying, 'We counted ethnic [redacted] as humans and turned the ethnic [redacted] from animals to humans', which was applauded by many of the participants in the meeting (Muluken, July 16, 2021). This video also shows that ethnic-based hate speeches, such as animalization, can be made publicly, with adequate level of support from ingroups.

Using cockroaches to portray others, which is the hallmark of the genocide in Rwanda (Ilibagiza and Erwin, 2006), is also seen during the conflict in the North, as seen in the excerpt below:



The excerpt above can be translated as 'it is the weak/coward who ceases to remove cockroaches from his home; we will soon make Tigray free from any cockroach and allow Tigray to be only for Tigrayans; Tigray will prevail'. This expression is made during the Conflict in the North, especially during the federal army and its allies controlled most Tigray region. In essence, the expression entails not only animalization but also its action component—eradicating target. As the source is more likely to be referring to the Ethiopian National Defense Force and its allies, specific ethnic group is not directly implicated in the

label. In the expression cockroaches are used to imply the ‘invasion’ of land by alien force. It, hence, urged the ‘brave men’ to discharge their responsibility in eradicating the alien force, while implicitly labeling defiant ingroups as ‘weak.’

While the above animalization excerpts are posted by unknown social media users, the data also showed high-profile names such as activists and politicians engaged in animalization of their targets, as the following excerpt shows:

"ከብቶች እየመሩን ፣ በከብቶች ጎዳና
እበት ፋንዲያ እንጂ ፣ የምን ብልጽግና !!"

The above excerpt from a high-profile self-claimed analytical journalist says ‘we are led by herds, so (expect) no prosperity but animals’ dung.’ The excerpt, which assumes the current Ethiopian leaders as herds and their ruling way as herds’ road, is framed in the context of discussion over a political issue that was happening in Ethiopia. The quote apparently targets the Prosperity Party (the leading political party in Ethiopia at the moment) which is repeatedly portrayed by its critiques to be unfairly ‘possessed’ or controlled by ethnic Oromo politicians. To crystalize this animalization expression, the above Tweet has obtained the following reply:

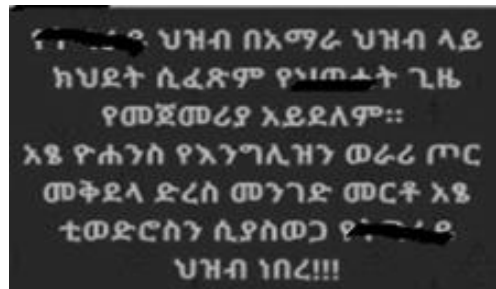
it's actually disgrace to the animals. Those ppl don't have any creature to match. They are alian.

This Tweet, in response to the previous one, not only supports its precursor but also adds more tone to the animalization expression. It claims that the group referred does not reach the level of animals; or that animals are superior to the despised target group under the discussion.

Animals are frequently used in animalization expressions as behavioral analogies, in which the group being discussed is assumed to lack human traits or to possess the animal trait in reference. In most of the expressions noted, therefore, comparison is made between the animal and the target referred, mainly in terms of behavior, trait or ability. As a result, while some put the target as equal to the animal, others even consider the target group to be less or worse than the animal used in the reference.

The data also shows such expressions partly feeds on the offline stereotypes and political disagreements. For example, the word ‘daylight hyena’ was used in recent years in Ethiopia mainly to refer to the leaders and accomplices of the Tigray People Liberation Front (TPLF) that were accused by the Ethiopian Federal Government of the lethal grenade attack at Masqal Square on people gathered there to support the PM, Abiy Ahmed, on June 15, 2018 (iMAGE eTHIOPIS, 2018). This expression was used again to refer to the same target, presumably attacking the Ethiopian National Defense Forces at midnight while they were asleep. Apparently, ‘Daylight hyena’ metaphorically would mean the target has a killing nature, poor vision, is distressed, is extremely dangerous, has lost all its senses out of sheer greed, and has an attacking propensity that comes out of confusion and helplessness.

The dataset also shows that, in the context of the conflict in the North, the daylight hyena expressions against TPLF/Tigrayans were also augmented with equivalent expressions such as ‘termites’ (implies working furtively to slowly kill its prey), chameleon (i.e., unreliable), and ‘serpent’ (being poisonous or deadly), all of which contributed to the rhetoric of describing members of the ethnic Tigrayans as conspiring and spying for TPLF, as the post below which is taken from Facebook shows.



The expression in the above post, which is in Amharic, is translated as ‘this is not the first time ethnic [redacted] have betrayed ethnic Amharas. During the era of King Yohannes, the [redacted] people were instrumental in collaborating with British colonizers in attacking King Tewodros.’ In this excerpt, King Yohannes is believed to be a descendant of ethnic Tigrayan and their representative, while King Tewodros is portrayed as a descendant of ethnic Amhara and their representative. This expression also targets the entire ethnic groups than elites or politicians. In essence, it recollects accusatory memories and portrays its target as unreliable, betrayals, and bandits, epitomizing the labels ‘termite’, chameleon, and daylight hyenas.

As seen above, animalization expression include portrayal of ethnic groups as ‘cattle’ and ‘herd’ both of which, in the context in which they are found, presumably connote lacking the

mental agency to think rationally. This is seen in expressions found in the dataset that do not directly use animal names but hold equivalent meanings, such as: ‘human-like creatures’, ‘have no mind to think with’, ‘groups who have not finished evolutionary processes’, and ‘groups lacking criteria of humanness’. The use of donkey to derogate target group metaphorically would imply that the target is uncritical as well as not deserving of respect whatever supportive or helpful it is.

While much of the use of animals as an analogy assumes the referred groups have the same undesirable traits as animals, there are also ways in which the human groups are portrayed to be less than animals, as in the post below, in which an unidentified person is allegedly burned alive by a man who appeared to be wearing soldier's clothing.



The caption in the above post assumes the use of ‘animals’ as an insult to humans is a misnomer and a disgrace to the animals, as animals have better conduct than humans. The tweet on page 81, in English, also shows an extreme version of animalization expressions of hate speech- less than animal. Apparently, at least in the above post, the hate monger portrayed its target as less-than-animals presumably due to the highly criminal or immoral act which the target is believed to have committed. Indeed, the dataset contains some more excerpts that shows ‘less than animals’ expressions used to describe ethnic groups, mainly because of the alleged immoral or criminal acts committed by their members. This shows hate speech is generalized from individual actions to the group to which they belong. Table 6 shows some ‘less-than-animal’ expressions found in the dataset:

Table 6 Less-than-Animal Expressions

S/N	Excerpts
1.	They did what is expected from them; they are lower than animals. What is surprising would have been if they didn't do it.
2.	It is difficult to label [redacted] as beasts, let alone as humans! a hyena, lion and Tigers do not eat the flesh of their own kind [they are not cannibals]
3.	It is surprising to see humans displaying traits less than animals. zebras, Giraffe, and other animals live in a field without fight! But it is painful to see these less than animal beings!
4.	In general they can be labeled as 'worse than beasts', because wild animals like bears never touch a dead body once they know it is lifeless!)

The less-than-animal expressions seen in Table 6 show the labeler has ' reasons' for describing a group as they see some offensive actions committed by the group the led them to the labeling. For example, the expression in number four indicates that animals do not touch dead bodies, presumably implying that the target has done something wrong to a dead body. The less-than-animal expressions also show it is not the actual perpetrators who are insulted but also the ethnic group they belong to. Most often, the less-than-animal and worse-than-beast expressions take moral issues, cruelty, and intellectual capacity while comparing the human group with the animals used in the expressions.

4.1.1.2 *Objectification*

In addition to animalizing expressions, the dehumanization data also includes expressions that portray outgroups as some kinds of undesirable, harmful, or useless objects, such as stones, animal waste, piles of litter, weeds, cancer, and placenta. I named the conceptual territory of describing human beings as such *Objectification*. Table 7 shows sample objectification expressions taken from the dataset and the contexts they are framed in.

Table 7 Objectification Expressions

Sample Excerpt	Immediate Context
These are stones who stood near other stones' (against Qeerro)	Ireecha celebration in Addis Ababa
Congratulations, Amhara is freed from cancers' (against TPLF)	The conflict in the North

As Table 7 shows, objectification of a group, as it is for animalization, connotes certain metaphoric meaning (i.e., why such a name is given) as well as entails what is expected from ingroups (i.e., taking actions such as removing cancer and weeds). An illustration of

objectification is a live transmission aired on Ethiopian Broadcasting Corporation in which a high-profile scholar portrayed TPLF as follows:

we have to think about what to do with TPLF. Cancer can be removed by operation as well as chemical and radiation. It is in these methods that a cancer patient can be treated. If one cancer cell survives, it will replicate and kill the person. **TPLF is a cancer**. As the proverb ‘eating fish needs techniques’ goes, we must have a program to **systematically and methodically eradicate all** woyanes.

(source: TghatMedia, 3 January 2022; bold to emphasize)

This highly toxic expression is made public by the elite presumably because TPLF was officially labeled as terrorist organization by the Ethiopian Federal Government. The observation made on the dialogue that went on among users while reacting to this video shows that many Tigrayans felt offended by these expressions to the extent that some say, ‘We all are TPLF.’ In addition, this highly genocidal call went viral mainly due to the intentional ‘sharing’ action on Tigrayans for whatever reason they have.

During the conflict in the North, high profile name such as politicians and public figures are also noticed sharing hate speeches in which they labeled their target as weeds (enboch⁴), cancer, and piles of dirt, and implicated the type of measure to be taken (removing or eliminating the target). The objectification expression, hence, implies the description of target as such entails some kind of action component.

Groups are also seen portrayed as waste/byproduct of animals and placenta. For example, while referring the TPLF, a public relation officer of a government ministry used the expression while addressing an interview with Ethiopian Press Agency⁵, “ኢትዮጵያ የእንግዳ ልጅን አየቀበረች ነው””, which could be translated as ‘Ethiopia is burying its **placenta**’ (bold mine). The label placenta apparently means ‘pseudo-offspring’ and presumably implies being useless and denotes TPLF which is believed to have betrayed Ethiopia, in the context of the conflict in the North.

Objectification is also manifested in how haters treat custodies and the dead bodies of their victims, such as the alleged leaving dead bodies of target groups to be meal for beasts,

⁴ A local name to a fast-growing invasive weeds, scientifically called Water Hyacinth

⁵ <https://www.facebook.com/armofethiopia/posts/677902023656743>

dragging human dead body on ground, and burning dead bodies or those alive as seen on page 83.

4.1.1.3 *Demonization*

Another manifestation of dehumanization is evident when a hate monger shared a video in which she portrayed ethnic Amharas as ‘sons of witches’, ‘Debtera⁶’, and magicians, in the context of the conflict in the North (Tamirat, December 21, 2021). This and many more social media contents in the dataset are identified to possess expressions that portray a group as Demon/Satan, its worshipers or collaborators such as magician, witches, or deadly beasts symbolically to mean having the propensity and lust to shed and /or drink humans’ blood. These expressions collectively make up the third subcategory of dehumanization, which I named ‘*demonization*’. Demons, monster, messengers of the dark, cursed groups, decedents of nonhuman cursed race, Debtera, blood sucker, cruel and savage, and deadly serpents are some of the terms and phrases identified from the dataset that describe targets as demon or its collaborator. Table-8 contains quotes taken from the dataset which are used in various contexts to demonize individuals and ethnic groups in Ethiopia.

Table 8 Demonizing Expressions

S/N	Excerpts	Contexts
1.	We have practically seen Satan/Lucifer in our very eyes	Killing of innocents during the conflict in the North
2.	I think of Oromia as a place where <i>monsters</i> who are coldblooded and merciless for infants and pregnant women are living in’.	The [alleged] killing of innocents in Wollega zone of Oromia region.

In addition to demonizing groups, Table 8 shows observable heinous acts committed are used as the reason for the hate speech. It also shows the fact that the entire region is negatively portrayed because of the alleged misdeeds of a few. Another illustration of demonization reads:

⁶a respected religious figure in Ethiopian Orthodox Tewahedo Churches who is well versed in aspects of the liturgy, and who sings hymns and dances for churchgoers. In the hate speech saga, yet, some use the word to mean witches.

You *Satan*, messengers of the prince of dark ...you killed us via the beasts [i.e., solders], you made us orphans, brutally killed innocents, demolished our holy sites with artilleries, killed our priests, raped our women, executed our youth, ruined our infrastructures, you are deadly serpents....

The above excerpt is a comment given to a video ‘በሐራ ድሬሮቃ ግንባር የሚገኙ የፋኖ አባላት ጅብድ’ broadcasted on Amhara Communications on September 1, 2021 in the context of the conflict in the North.

In my analyses, demonization can also take the forms of barbarism such as brutal killings. The following excerpt is taken from a discussion of three academic elites in which a professor and PhD holders participated. While referring to the Geda system and ownership of Addis Ababa, a discussant who is a self-claimed historian is quoted saying the following:

Before the ██████ invaded, scholars had been living in Shewa province. The ██████ clan brought with them acts of barbarism like cutting genitalia when they arrived in Ethiopia's Shewa province. There is also proof that the Abune Abraham monastery's monks and residents had their genitalia cut by the ██████. They used to emasculate others. I think that their arrival in Ethiopia caused the spiritual decline of the people (Tamiru, September 2, 2020).

(ethnic name shaded)

This kind of portrayal of Oromos as ‘natural killers’ in history is also observed from another social media user who described Geda system as inherently brutal, as the caption of the post below, which is a typical Aba-geda of Oromo, shows.



The caption in the above picture which is in Amharic reads: ‘this man is Aba-Geda⁷, the material in his forehead is male genital that shows historically they were cutting others genitals.’

⁷ a highly respected cultural leader in the Geda system of Oromos

The data also shows presence of ethnoreligious demonization. This is more evident as the Ethiopian Orthodox Church Priests are portrayed as magicians (debtera in hate speech saga) in such a way that the expression describes both the church and ethnic Amhara as being agents of demon-witches.

In many instances in the dataset, Qeero, Fano, TPLF, ENDF, OLA, Debtera, and King Minilik II are frequently portrayed as demon and its associates. As we have seen the ‘less than animal’ description, the dataset has also expressions which portrayed outgroups as ‘worse than demon’. An illustration to ‘worse than demon’ description of others is seen when a political leader, in the context of the conflict in the North, portrayed people living in Tigray as ‘worse than Diabilos’ (i.e, an equivalent term for demon in Amharic). The source of the message says ‘Satan doesn’t have more devilish technique than these people’. This message is broadcasted on ‘ቀይ መስመር - የድሬዳዋ ጉዞ ወደ ቀደመ ስሟ’ ‘The Red line’ program on Ethiopian Broadcasting Corroboration, a television platform owed by the government, dated December 20, 2021 (Abay, December 20, 2021). Same level of ‘worse than demon’ expression was echoed by a government body who, in the context of the conflict in the North, portrayed TPLF fighters as ‘worse-than-demon’, monsters, weeds, inherently and structurally evil collaborators, and groups unknown in or far from the animal kingdom for their heinous act.’(Prosperity Party - ብልፅግና, December 20, 2021).

Three of the subcategories presented above: animalization, objectification, and demonization categories have something in common: they deny the full human status from their target and assigned other non-human attributes to them and hence are manifestations of dehumanization.

4.1.2 Enemification: they are enemies and threats to us

The dataset includes materials that portray out groups as enemies or threats to the in-group, attributing to them traits or deeds such as a proclivity for economic exploitation, land grabbing, oppression, marginalization, and cultural genocide, and/or killing and destroying in-group members—mostly ancestors. Because the common theme threaded throughout such expression is seeing others as threats and enemies, I set conceptual boundaries to such expressions under the label *Enemification*. The expressions either directly call the target enemies as enemies, or portray them as such without using the word enemy. Table 9 shows some of the excerpts that illustrate enemification:

Table 9 Enemification Expressions

S/N	Excerpts
1.	We shouldn't cry in vain anymore. We rather strengthen ourselves/unity and commitment to react to these animals. <i>We now know they are our enemies.</i>
2.	Only Amhara especially Gonderite and Isayas Afewerqi are the major enemies to Tigrayans
3.	The usual interest of Amhara is to assimilate other ethnic groups , diminishing their identities, looting their resources and history, and finally become a sole leader of Ethiopia.

NB: bold to emphasize

The expressions in the table implies that the labelers have repeatedly tolerated something bad from the target, have a preferred group to be labeled as enemy, and that infers stable negative trait that qualified them as enemy. In some of such enemification expressions in the dataset, the tone is strong as seen in using adjectives like ‘historical’, ‘inherent’ or ‘eternal’ to the label enemy. The enemification expression manifest itself in different themes, as described below:

4.1.2.1 Land grabbers, assimilators, and settlers

In the first subcategory, enemification employs accusatory narrations and conspiracies to describe its target as *assimilators, land grabbers, and invaders*, and hence as a *potential threat* to the present and the future life of ingroups. More importantly, economy, culture and land are mainly implicated in enemification expressions, as the following two posts show.



The post in the left, which later on is found to be photo-shopped, portrays the Oromos (or its politicians) as if they seized the land of Southern Nations and Nationalities People (SNNPR). The source of the post put a wrong caption to an Oromo politician standing in Borena national park and described the scene as if it is in Nech-Sar National Park which is under the SNNPR. This picture also reminds us that hate speech is partly achieved through disinformation by twisting realities. The Amharic caption has the following meaning in

English, ‘not only the earth, the sky is also ours’ which implies having incurable lust to grab everything they come across. The post on the right, on the other hand, is an instance of accusation against ethnic Amhara by portraying them as grabbing Oromos’ land in the name of establishing Orthodox Churches, and calls for ethnic Oromos to be cautious of what he named the invading forces ethnic Amharas. The actual Amharic caption in this post can be translated as:

Their own priest is exposing how OrthoAmharas* are grabbing Oromos Land deceiving the local people into believing that an Arc has descended from the sky of God (to convince them leave the land for constructing Churches). My Oromos, please wake up as the enemy is taking your land under the guise of religion.

(* to mean Orthodox Christian Amharas)

This post, on the left also shows the presence of hate speech that targets both religion and ethnic group (i.e., ethnoreligious hate speech), as well illustrated in the excerpt below, which is taken from a different context:

The Orthodox Church anointed King Minilik and collaborated with him in enslaving other ethnic groups and empowered themselves with the arms they bought from Italy, UK, and France. Amhara Orthodox church followers, the church, and Minilik II should be held accountable for the genocide their committed against other ethnic groups. It is difficult to differentiate/separate King Minilik II from Orthodox Christian.

The excerpt above accuses an individual, an ethnic group, and the Orthodox Church as if they collaborated in genocide and slavery. It also states Minilik II and the Orthodox Church are inseparable before history. The above two enemification excerpts also show how hate speech is embedded within politics and religion. The second post, for example, are observed seasonally and mainly targets the Amhara groups who are frequently portrayed as proponent and guardians of the Ethiopian Orthodox Church. The data also shows enemification was evident in the context of national election, and that the issue of land grabbing prompts labelers to portray targets as enemies, as seen in the post below:



የጠላቶቻችን ዕቅድ እና አላማ የአሮሞ ህዝብ በሂደት ጠፍቶ፣ እንደዚህ ሰፋሪዎች በአሮሚያ ላይ የሚመራረጡበትን አገር መፍጠር ነው። ይህን አሮሞን ከመራቱ ነቅሎ፣ በእኛ አገር ላይ እነርሱ ባለአገር የመሆን ስልት "ንቁ ዜጋ! ምቹ ሀገር!" ይሉታል።

የኢዜማ ለአዲስ አበባ ከተማ አስተዳደር ምክር ቤት
የልደታ ክ/ከተማ ዕጩዎች
ኢዜማን ይግረጡ!

ንቁ ዜጋ!
ምቹ ሀገር!



አትተክሌ በቀለ ደመረ
በጦርዳ ዐፋ የኢዜማ
የህ/ኅ/ም/ቤተ-ዕጩ

ግርማ
2013

ራዕዮተኛ ጠቅላይ ሚኒስትር ገብረመስቀል ገብረመስቀል



ሚሳኒ ነገው የግርማ ምልክት

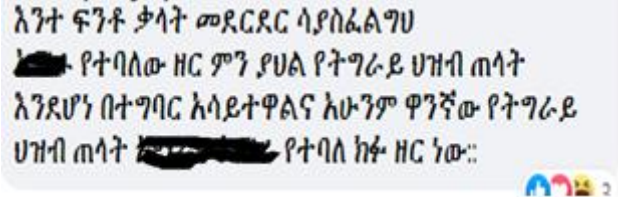


[https://twitter.com/\[REDACTED\]/status/1405217512715788288](https://twitter.com/[REDACTED]/status/1405217512715788288)

In the below post its captioned reads ‘Our **adversary**'s intention is to systematically eradicate ethnic Oromos and give the land to the settlers. They called the technique of evicting the aboriginal Oromos and replacing by the invaders 'alerting citizen, conducive country.’ The essence of this post is congruent with many excerpts in the dataset that portray outgroups as settlers and land grabbers. In the same token, Tigrayans are not immune to the accusation for having incurable desire to seize or invade others land, as seen in the following deleted post from Facebook: ‘Not only TPLF, but also the entire Tigrayans want to grab the land of others; they are expansionist in their approach....’

4.1.2.2 Unreliable Historical Enemies

The second subtheme found under enemification is the portrayal of targets as historical enemies and unreliable partner with the tendency of betraying others, as seen in the excerpt below which directly portrays Amharas as historical enemies of Tigrayans and cursed races. The post is a comment given to another post titled: [REDACTED] ብሔርተኞች የማይሸር የአማራ ጥላቻ’ (to mean incurable hatred of Tigrayan ethnonationalists against Amhara) by a high-profile person from ethnic Amhara side.



The excerpt above can be translated as ‘In simple language we have practically seen how ethnic [REDACTED] are enemies to Tigray people. Currently, the cursed people of [REDACTED] are typical enemies to Tigrayans.’

In addition to describe target groups as enemies, the excerpt shows hate speech could surface in response to another hate speech targeting own group. In essence this kind of expression could be retaliatory ones given to in a ‘tit-for-tat manner. It also ranks the target as ‘first lever enemy’, calls them as cursed race, and claims having reason for these labels.

Such portrayal of outgroups as enemy is also seen from high-profile politicians. For example, a top level regional leader is caught making a highly incendiary and arguably genocide-inciting remark, which is broadcasted on government owned television on July 25, 2021: ‘ከዚህ በኋላ እረፍት የለንም፤ ይህ ህዝብ የመላው ኢትዮጵያ ጠላት ነው፤ ይህ ህዝብ የአሮሞ ጠላት ነው፤ ይህ ህዝብ የአፋሩ ጠላት ነው፤ የጋምቤላ ጠላት ነው፤ የሶማሌው ጠላት ነው።’ which is translated as ‘*We will not take rest until we have annihilated this enemy. These people are enemies of the whole of Ethiopians; they are enemies of the Oromos, the Afar, the Gambella, and the Somalis*’ (Lilo Media ሊሎ ሚዲያ, July 25, 2021) (italics and bold mine).

The dataset also has expressions of historical narrations in which ancestors of target groups had been murdered in-groups, while reminding that their descendants are of the same page in relation with in-groups. This is portrayal of others as having killing and destroying culture. Nonetheless, there are evidences that show presence of role reversal or reprisal that occurs in the accusation or enemification, whereby the group that was accused in one context becomes the accuser in a different context or time.

4.1.3 Devaluing: they are uncivilized, unlearning, cheap, or burden

The dataset contains a lot of contents in which outgroups are described as burden and cheap people (e.g., coolly or slaves in history, beggars, commercial sex workers, and hunger strike), and uncivilized or backwards (e.g., unsanitary and unlearning) in ways that evoke contempt and disgust. In this study, the conceptual representations that are observed in these kinds of expressions are named *devaluation*. Sample devaluing expressions are described in Table 10:

Table 10 Devaluation Expressions

S/N	Excerpt
1.	Amhara Farmers are better off than their academician; they are morons even if educated
2.	They don't have any good history except producing louse. The Amharas brought them from Sudan as laborers while constructing Castle of Fasil
3.	██████ means children of prostitutes; they are leprosy hit people
4.	Don't judge him as animals have no religion; they are out of control when their belly is full let alone getting such political power

As Table 10 shows, groups are described as unlearning entities, cheap and burden, belong to lower social class, and religiousness. Further scrutiny on the dataset shows, devaluation type of hate speech is manifested in four subcategories:

4.1.3.1 *Cheap*

Some of the hate speech expressions portray a group as slaves, settlers, burdens, or having some physical defect. The following quotes from the dataset illustrate this theme: ‘The word Tigre came from the Amharic root word ተግራ (perhaps means 'I engage in suffering due to hard work' or ‘under my foot’) which implies the forefathers of Tigrayans were brought by the Axumite Amhara from Yemen as their slaves’, ‘██████ are the same in the past and today too; they make a living by begging and working as commercial sex workers’, and ‘██████ are ‘leprosy hit people, coolly, settlers’. I encapsulated such devaluing expressions which assign its target the lowest social hierarchy which is detested and despised by others under the subtheme ‘*Cheap*’. It seems to me that some of the expressions appear to be targeting to irritate their target.

4.1.3.2 *Backwards*

This subcategory includes portraying the target group as uncivilized, inherently unwise/fools, and unlearning. Labels/adjectives such as ‘nomads’, ‘uncircumcised’, ‘they live the stone-age life’, ‘not able to cook and eat’, having incurable lower status, and ‘still follow cows’ tail’ are bearers of this subtheme found in the dataset. The backward subtheme is illustrated by excerpts from Facebook comments like ‘Amhara is living like the apes in the mountains’, and ‘you will never show progress whatever effort put on you.’ Another illustrator for this subtheme came from a speech by a critique of some ethnic Oromo nationalist forces and a frequently observed commentator on Ethiopian history, who said the following, ‘██████ are stupid, lazy people; they don't like working, and have no skill in their entire history’ (ተክላይ, November 10, 2022).

Two more illustrations for this theme are presented below: The post on the left below is a mockery against ethnic Oromos in which the source portrayed the Irrecha celebration in Addis Ababa to produce litter-presumably disgracing the people who celebrated it. The post on the right pictured video footage that is taken from Amhara Television. In the video, barefooted villagers are interrogated by host journalist working for Amhara Broadcasting Corporation regarding why they don't wear shoes to which they responded that they are averse to shoes. This video later on fell in the hands of a hate monger (a poacher in this study) who used that as evidence to describe the entire ethnic Amhara as being archaic and resistant to wearing shoes (Elias, March 7, 2022).



4.1.3.3 History-Less

The third subtheme identified under devaluation is portrayal of a group *history-less* which includes depiction of groups ‘not Ethiopians by origin’, ‘cheaters of others history’, and ‘have dirty history’ or ‘have no good history’. An illustration to this is the following excerpt: ‘we shouldn’t have expected better things from [redacted] leading the country is not their gift’. The subtheme **history-less** is also found to be accompanied by fierce competition over history and resource ownership, such as lands (e.g., Finfine/ Berera), languages (i.e. owner of Amharic/Geez letter), historical narrations about bravery and heroism, and artifacts such as Axume Obelisk and Lalibela Rock-hewn Church.

4.1.3.4 Aberration and Conscienceless

In the fourth subtheme, *aberration*, target groups are assigned undesirable psychological traits or behavioral tendencies such as being inherently *betrayals and traitors*, and conscienceless as seen in the following comments ‘... you son of traitors’, which is supported by this post which accuses Tigrayans for being thieves as seen in the Facebook post below written in Amharic.

የተገኘው ህዝብ ሌባ ነው ገንዳ ነው እስመሣይ ነው ቅማላም
ገንጋራም ህዝብ ነው እምልበት ዋናው ምክነያት የሚከተሉት ናቸው ::
1.127 እመት ስሙ ስን ስመሠለፍ ዘርፈዋል ፣ ከብረዋል

The above excerpt in Amharic can be translated as ‘It is because they collaborated with [redacted] in looting the country for 27 years that I call the entire [redacted] thieves, traitors, and shrewd.’

The aberration subtheme is also manifested in looking at the target at the lower ebb of the moral hierarchy such as lacking religion, conscience or moral compass (*Consciences-less*), as seen in the following excerpt taken from a video transcript:

Had it not been for the Amharas and Tigrayans, the [redacted] would have been remained pagan! They knew only Wakefeta⁸, Irecha and worshipping trees! It is Amharas who taught Oromos Qine-Mahilet⁹, hymn and the religion.

4.1.4 Attacking and Repelling

The fourth category of hate speeches, attacking and repelling, encompasses expressions which call for some kind of action or omission so as to punish or hurt the target group. This core category includes calls made to either to attack, evict, discriminate against, withhold help, or maintaining distance from target outgroups. In essence this category of hate speech is the action component of all the three types of hate speeches presented earlier.

The following quote illustrate part of the conceptual spaces of this category of hate speech:

Other ethnic groups should stand firm by this ethnic group (Gumuz) because you will be bullied/killed next by this same perpetrators. And we all must wipe out with broom the Amhara from everywhere they live! They don’t deserve to live in every region of Ethiopia, except theirs- I swear to God!

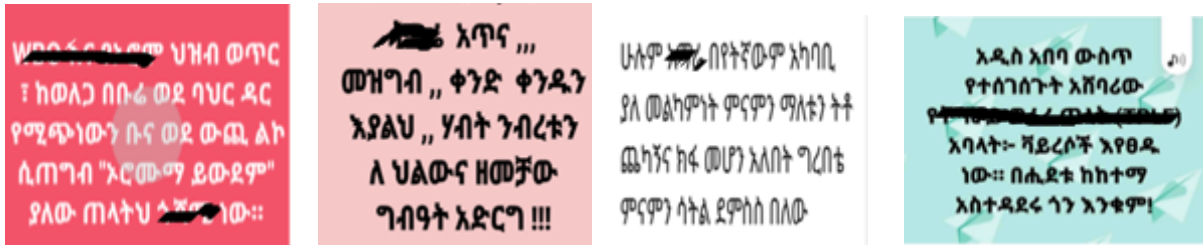
This call has lots of elements to be noted. It calls others to be united against the target; tries to evoke survival concern; calls actions to remove the target which it implicitly portrays as dirty (that is what broom is used for). This category of hate speech has the following subcategories:

⁸ An ethnic religion indigenous to the Oromo people

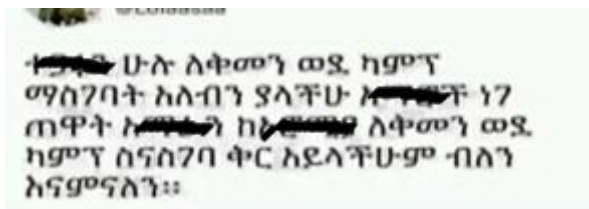
⁹ Religious song and dance performed in the Orthodox Church

4.1.4.1 Direct Call for Attack

In direct call for attack, supporters are told to attack targets, as seen in the following four social media posts:

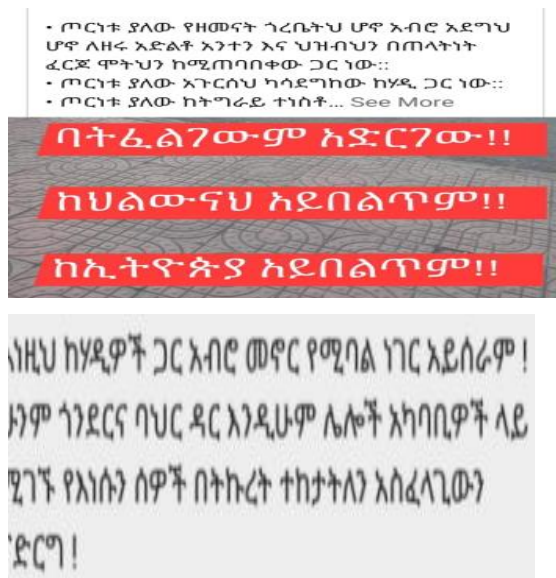


The above four posts have various issues to be noted. In the first post on the left, the source directly described an out-group as enemies who got affluent by exploiting and victimizing in-groups. The post also calls for armed groups of ethnic Oromos (‘OLA’) and the entire Oromo people to fight back [redacted] people. The word ‘ወጥር’ which mean ‘be firm and committed’ imply the commitment needed to fight back the labeled enemies. In the second post above, the call starts with the need to identify and **register** potential members of the out-groups, to be attacked and looted later on. This call is made during the conflict in the North and possesses the issue of existential threat as the reason to attack the target. In the third post, **emotional hardening**, which excludes the targets from the moral circle, is called for so as to effectively destroy the target. The Amharic word ‘ደምስሰ’ literally means ‘eradicate’ and is a word that shows the call is made to conduct genocide against the target. The call for emotional hardening is also evident in the following quote taken in the context of discussion on Tigrayans, and targets TPLF supporters: አረመኔን በአረመኔ መንገድ ብቻ ነው የምናሸንፈው (which means in order to be victorious, we must respond cruelly the way they did). This excerpt reminds members the need for emotional hardening against the target group. In the fourth post above, the message uses the word ‘virus’ and that supporters should join hands to facilitate the cleaning of the virus which is ongoing at the moment. In the tweet below, a social media user records others call for attack and *backfires same gun* at the source.



This Amharic tweet can be translated as ‘Those of you A [redacted] as who said you must put [redacted] ns in concentration camp, do you feel bad if we do same to you?’ This post has two issues to note: the first is the idea of concentration camp, an important context to carryout

genocide as seen in Nazi Germany against the Jews (Morrock, 2010). It has to be noted that although the idea of placing some people in concentration camps was suggested by some during the conflict in the North, primarily for security and safety reasons, some others feared it might encourage or facilitate genocide. Second is hate speech are responded with another hate speech (that could be justified in a tit-for-tat manner), in a way that enhance polarization and rifts between groups. These kinds of calls for attack, such as ‘clean them’ expressions, were rampant during the Tigray conflict. What makes the call scarier is such messages were also taken as a response to *survival concern* and taking ‘proper’ measure was seen as the sole option left to in-groups to avoid what they call imminent danger. The fact that call-for-action takes survival threat as the main motive has huge impact on the attitude and action of the audience (to accept the call to escape the perceived threat or demise) and is termed in psychological literature as accusation-in- a mirror (Marcus, 2012).



In the above two posts, the need to attack out-groups indiscriminately is communicated on the ground that this is necessary for survival of in-groups. What makes these posts more dangerous is that they are communicated by persons who are respected figures in their community. There is also role of interpretation in which a hate speech is extrapolated as call for measure. In Picture below, for example, demonstrators are seen holding a banner which reads ‘let Oromuma be destroyed’, to which an apparently ethnic Oromo social media user felt offended as he interpreted the content and assigned more meaning beyond the actual content in the banner. This post reminds that hate speech could employ twisting reality and misinterpreting others message, which will be presented in section 4.2 and 4.3.

- እርሞ በግፍ ሲገደል የመንግሥትን ህግ የማሰከበር ሂደት እንደግፋለን፣ከወያኔ በኋላ ወደ እርሞ እንዘምታለን የነሱ እዳው ገብሰ ነው፣እርሙማ ይውደም፣መንጋ ...
See More



Innocent members of outgroups are also targeted for retaliation in the call for action made (scapegoating), as seen in the following excerpt which is in response to the news circulating in the social media about the killing of ethnic Amhara students in Universities found in Oromia region: ‘በመላው አማራ ክልል የሚማሩ የ [redacted] ዩኒቨርሲቲ ተማሪዎችን በመግደል የአፀፋ መልስ መስጠት ያስፈልጋል :: አማራው መግደል ካልጀመረ እነሱም ከመግደል እና ከማሳደድ ወደኋላ አይሉም’, which is translated as ‘we must retaliate by killing [redacted] students who are learning in Universities found in Amhara region. Unless we start killing them back, Amharas will continue suffering such killings.’ This and similar other call for attack made on social media show presence of intention to take retaliatory measures (at least as ritualized aggression form) in response to a previously induced harm.

In a call to action, the sources can also engage in what I termed as *ally formation and separation* in which the source calls different outgroups to join hands together against the target group, as illustrated in the following excerpt ‘Why is that Qeerros are not killed as they kill others? Where is our unity as Ethiopians? We must have one voice and each ethnic group should stand for the others.’ The excerpt presented on page 96 that calls allies to join in attacking Amhara is also an example for this theme.

4.1.4.2 Warning and Threatening

The second subtheme under call for action and ally formation is **warning and threatening** in which target groups’ survival and wellbeing are told to be conditional in such a way that unless they fulfill certain preconditions (i.e., *blacklisting*), they will be attacked, killed, deported etc., as seen in this excerpt, ‘ [redacted] are Chameleon; you will cry sooner or later’, and ‘unless the settlers of Minilik II keep quiet, they will be deported to their land.’

4.1.4.3 Cursing and Ill-thinking

The third subtheme identified is ‘*Cursing and Ill-thinking*’ in which the source expresses the desire to see the misfortune of the out-group and wish supernatural power could punish this

group, as seen in the following excerpts, ‘May the womb of ██████n mothers be infertile’(in the context of discussion over 2023 Olympic medal), ‘Even in modern times, ██████s worship trees... God's wrath be upon you’, and ‘They must be eradicated from the earth like dinosaurs’.

4.1.4.4 Call for Distancing and Repelling

The fourth subtheme identified in this category is call made to *distance and repel* out-groups from in-groups or the larger community such as withholding helps and avoiding befriending them or living with them. The dataset has three videos in which individuals call their ingroups to divorce if they have a spouse from another ethnic group. In a two-minute video an individual presumably from ethnic Amhara, is quoted saying the following:

Amhara women who marry or give birth to G████a men or Tigre-lousy men, May God make you infertile; may your womb produce worms and cancer. May your womb get incurably wounded for good if you marry Oromo or Tigre men! I love to see you suffering from that cancerous disease. If you marry and give birth to Tigres and Oromos, you produce enemies to Amharas.

The dataset also contains equivalent calls made by a high-profile name to Tigrayans to boycott Ethiopian owned business, and other social services including mosques and churches (Mereja TV, April 27, 2021); and thwarting helps coming from the target, as seen in this excerpt ‘we don’t need your help; you bastard settlers, we never sleep without paying you back for the blood you shed.’ This comment was made in response to a social media post that was likely made by someone of Amhara ethnicity who was coordinating help for Tigrayans who were supposedly suffering as a result of the Conflict.

4.1.4.5 Calls for and Expressions of Indifference

The fifth subtheme identified under this category is being ‘*Indifferent*’ to the pain and suffering of the out-groups-that is emotional hardening. This mainly implies expressing happiness when target groups get hurt or lose something valuable such as the death of respected members, or when the target experience misfortunes. The intention of the expressions in this subtheme is more likely to stick a stick in the target's wound, or to add salt to one’s wound. For example, the dataset included a post with mockery on internally displaced people who are apparently from ethnic Amhara, while another post celebrated the death of Hachalu Hundesa, presumably to irritate his admirers. By the same token,

‘outgroups’ are seen celebrating the death of a renowned professor whom they believe to have frequently disparaged their ethnic group. During the data collection, the post that describes his death received 2.9k likes, 893 shares, and 223 comments, of which 86% were favoring his death. This is a typical manifestation of emotional hardening. Some of the comments observed, for example, include: “Bravo”! Rest in Hell. Never back; and ‘one of the enemies of our people is illuminated’.

The following diagram summarizes the main hate speech categories and their components.

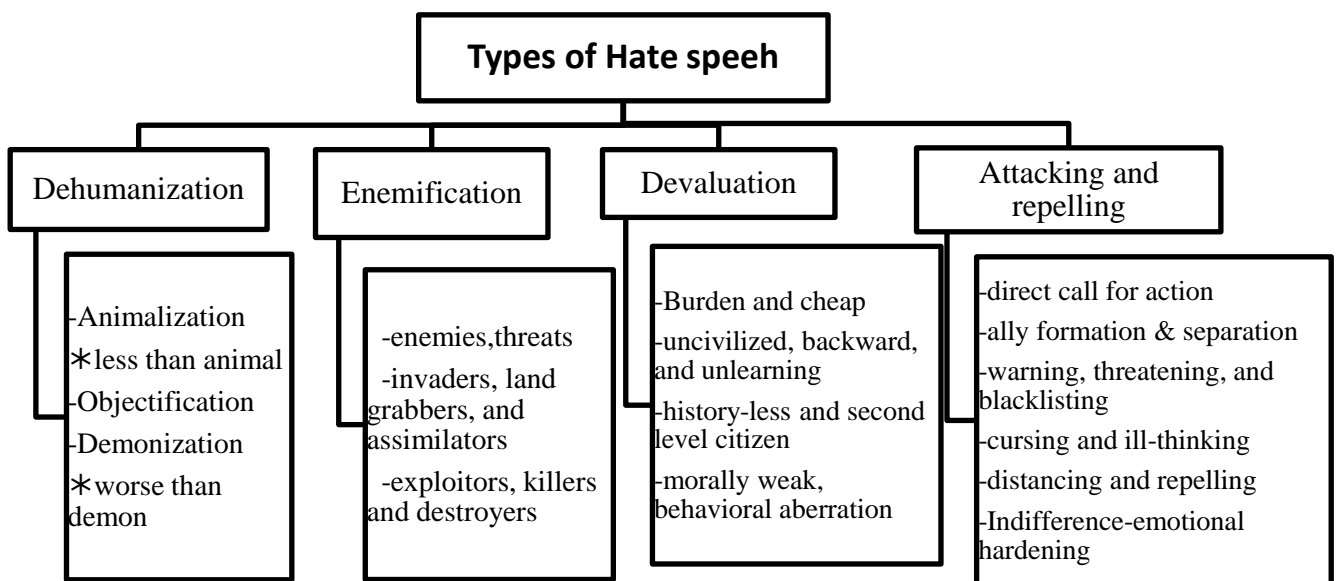


Figure 9 Summary of Types of Hate Speech

4.1.5 The Link among the Types of Hate Speeches

While each has its own conceptual territories, there is evidence that the four types are interrelated. The relationship among the three categories, for example, is more apparent when specific expressions are considered. Termite, for instance, implies animalization at surface level but ‘enemies from within’ metaphorically. In this sense, some expressions can have elements of more than one category, a surface level meaning and latent meaning. Devaluation and dehumanization can also be interrelated. For example, the ‘unlearning’ and ‘uncivilized group’ implies they lack the basic human features of rationality and learning.

The four types of hate speeches also vary along the description-prescription dimension. While description implies the expressions negatively portray the target group, the prescription

implies some actions are called for against the target. More importantly, the first three types of hate speech: the dehumanization, devaluation, and enemification are more of descriptions, while the call for attacking and repelling type is more of prescription-it calls for actions to be taken sooner or later. Nonetheless, even if the first three types of hate speeches are more of descriptions, they also have prescription elements embedded in them. For example, expressions such as snake, chameleon, Satan, enemies, and daylight hyena metaphorically shows certain negative behavioral traits assigned to the target and hence evoke emotions like anger, fear, contempt, and disgust. Such negative portrayal also implies taking some actions against the target group are necessary or justifiable. For instance, if a group is called weeds or cancers, there is a need to eradicate them. In that sense, the description (such as weeds, cancer), entails eradicating them is acceptable for one's safety and wellbeing. By same token, describing a target as historical and inherent enemies or as invader and assimilator requires some actions to avoid the 'imminent danger'. That is to mean, such expressions merely show what to do next if the target is described as such (the first three types of hate speech). Call for action and repelling type of hate speech is, hence, the action component while the first three are more of the cognition and feeling component of hate speech. Table 11 contains excerpts that show the link among the four types of hate speeches:

Table 11 Links among Types of Hate Speeches

S/N	Excerpt	Possible links
1.	These people are betrayals; they must be evicted from all the regions	Animalization with Calls for attacking
2.	We must take offensive action than merely defending ourselves. Otherwise our survival would be at the mercy of these beasts	Enemification with attacking and repelling
3.	If the narrative is true, they even deserve to beheaded, let alone worrying for their breast. They are piles of morons and deserve to be eradicated like Dinosaurs.	Devaluation and attacking and repelling
4.	As the pagans don't know about history, artifacts, and religion, they are ready to destroy them when they get the change to rule	Devaluation and enemification
5.	Begging and commercial sex work are a means of subsistence to Amharas They are animals	Devaluation and animalization/devaluation
6.	Animals have no religion	Dehumanization and

The links indicated in the above table shows hate speeches expressions can include two or more types of hate speeches. Diagrammatically, the relationships among the four types of hate speeches can be portrayed as seen in Figure 10:

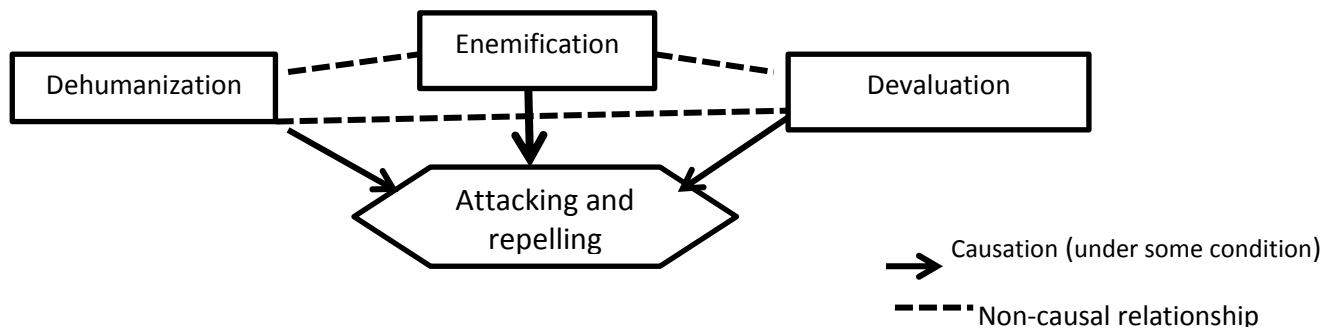


Figure 10 Proposed Links among the Types of Hate Speech

Figure 10 shows, while the first three types of hate speeches are linked to one another in non-causal manner, they are causally linked to the attacking and repelling category.

4.1.6 Contexts for Hate Speech

In the pursuit of understanding the nature and dynamic of ethnic based hate speech in Ethiopia, it has become very helpful and necessary to understand the contexts in which the hate speeches are framed. With regard to its timeframe, this dissertation heavily relies on data fetched during the political landscape after Abiy Ahmed came to power. This period is characterized by higher rates of ethnic based conflict, internal displacement, inflation, and tension among various political factions (European Institute of Peace, 2021). The period is also reported to be a time where hate speech and fake news have been exponentially increased, to the extent that the government responded developing a legal framework to suppress and punish hate speech and fake news. It is therefore important to understand the contexts in which the hate speeches are identified. These contexts are:

- i. **The Conflict in the North:** For example, the two year deadly civil war between the federal government and TPLF took place during the data collection period. This conflict has remarkably drawn what otherwise would be peace-loving citizens into hate speeches. It appears to me that conflict supportive narration and conspiracy theory bind them all the contexts. The conflict in the Northern part of Ethiopia created mutual accusation and helped hate speeches to be framed within ‘survival concern’. For example, while compiling this manuscript, Alice Wairimu Nderitu, Special Adviser on the Prevention of

Genocide said the following in his press release on 19 October 2022, in the context of the conflict in the North: ‘Targeting of civilians based on their ethnicity or perceived affiliation to the warring parties remains a key characteristic of the conflict and one that is exacerbated by horrifying levels of hate speech and incitement to violence.’

- ii. **Competition over History and Resources:** as seen in the data collected and analyzed in this study, Ethiopia’s nation-building is rife with accusations and claims of past atrocities and injustices from the time of Emperor Menelik’s southward expansion to recent political discourse since 2018. The majority of hate speeches analyzed in this dissertation is hence can be traced back to ‘Ethiopian History.’ More importantly, ownership of land and territory is highly implicated in the hate speeches analyzed.
- iii. **Low Level of Diversity Tolerance:** A few of the hate speeches expressions observed in this study are framed within ethnic markers, such as showing disrespect against others culture and tradition.
- iv. **Failure to uphold Law and Order:** a lot of the hate speech expressions are also framed in the context of criminal and heinous activities committed by members of a group, such as extrajudicial killing, identity based attacks, displacing citizens from their residential areas, publicly offending others, and committing any crime like corruption. These actions, I argue could have been prevented provided that the country had stronger law enforcing mechanisms or systems. Nonetheless, the failure to uphold law and order has lent ethnic groups to hate speeches, for the very reason of the wrongdoing of their members.

Elites, including politicians and academicians, engage in hate speech as well as lend their group to hate speech when their media appearance is unappealing. It appears, hence, that elites of a group strengthen the dominant stereotypes others already hold against the group. In addition, any insult made against elites is taken as an insult to the entire group, and hence attract ingroups to the hate speech circle.

Section 4.2 How Do Social Media Users React to Ethnic-based Hate Speech Posts?

(typology of users with regard to responding to hate speech posts)

The second research question addressed in this dissertation is about how social media users react to ethnic-based hate speeches that they come across online. The data for this research question is only comments given to hate speech posts than the primary post itself. It is apparent that a hate speech post could beget positive and negative reactions from a wide range of users. It has to be noted that users' comments might be directed at their members (in-group), to relevant outgroups, or to general social media users. After extracting the comments from selected hate speech posts, the analysis is made on 940 comments. The identified roles are further scrutinized so as to enrich the attributes of each category. Based on the analysis made, the followings are the roles social media users are playing while reacting to hate speech contents they come across online. It should be noted that even if five types of roles are identified from the comments users provide to hate speech posts, there are evidences that users play these roles in their overall social media usage (as posters than commenters).

4.2.1 Trolling

Table 12 Sample Reactions of Users to Hate Speech Posts

S/N	Excerpts
1	You are psychotic, donkey
2	May God demolish/punish you as He did against your relatives
3	Tigre introduced anti-Orthodox Tewahido Manifesto
4	Woyane, you are going to be vanished
5	████████ slaughter innocents because of the false narratives promoted by Tigrayans
6	Human can be created like this stupid woman
7	you are killed because of your narrow mindedness; you are stupid

The above excerpts in Table 12 have common element; they offend others. I labeled a category of users' reactions that characteristically offend or irritate others Trolling. This category of users' reaction to hate speech, therefore, comprise reactions characterized by insulting, accusing, and cursing others, disrespecting, mocking at, showing indifference etc. Trolls appear to be inherently social media polluters and on the business of offending their target (i.e., offend the poster, other individuals or group). Trolls can react to: a) the source of the hate speech post, b) the group the source of the post belongs to, or c) any third party not directly implicated by the hate speech post (i.e., another individual or group) as well as a combination of these.

Some trolls appear to be constantly annoying their targets, for virtually every reason they find. They are true haters or have the propensity to negate others on the platform. Trolls could also behave rudely when reacting to positive posts. That means they offend people even while responding to apparently neutral or positive social media posts. For example, examining the posts a troll shared for a solid month, nearly 85% of the posts possess the target ethnic group's name the troll appears to abuse, offend, or belittle.

Trolls are also seen to engage in excessive accusation, cursing, and ill-thinking as seen in the following comments, 'May God Shorten your life', 'May Allah Burn you!', 'May God destroy you like he did on your relatives', 'God has blinded your eyes and mind because you are a cruel and bad man; may He pay you more.' These comments clearly show that trolls are assuming a-just-world-view, *which* maintain that individuals and groups get (and should get) what they deserve, because the world is a just place (Lerner, 1980).

Two types of trolls are understood to exist: *the venom types and the parrots*. The venomous are full of malice or demeaning remarks. As the venom is secreted by a gland inside the animal, these users use toxic words while providing comment to express negative attitude towards the subject of their speech, as evidenced by the immediate context, the tone of their expression, and the consistency they displayed. Among the total of comments analyzed 44.6 % (420 of them) constitute the venom subtype. The venom subtype could spite their derogatory remark either on individuals or on their group. In the first, what we called *venomous-personal attack*, they insult or derogate individuals, yet without attacking their ethnic groups. Instances of venomous personal attack observed are directed at the physical characteristics of the individual (ugly, dirty), behavioral (gossiper, terrorist, devil), cognitive (moron), or a combination of these (e.g., 'sick person', psychotic), with implied motive to disgrace that person per se before the audience. The table below shows excerpts that show venomous personal attacks:

Table 13 Sample Venomous Personal Attack

S/N	Excerpts	Attacked part of the person
1.	You cursed man, both your eyes and mind are blinded	Cognitive and physical
2.	You Crossed-eye!	Physical
3.	You are very ignorant person	Cognitive
4.	You moron/deaf, you worship Trees	Spiritual and cognitive
5.	Kkkk, Stone/meat head	Cognitive
6.	አጋሰሰ (pig!)	Physical or behavioral

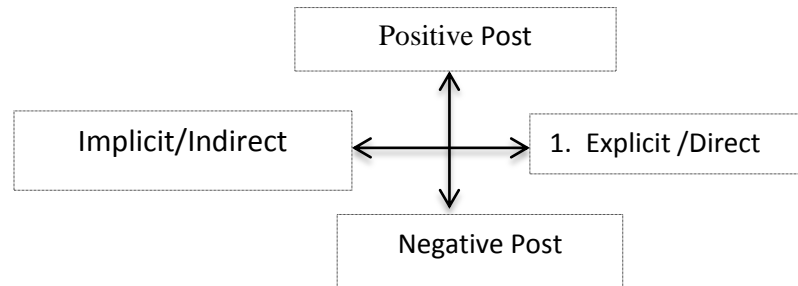
The venomous personal attacker could also attack two or more of the protected characteristics the target, such as ‘Muslim Ga■’ or with two different negative labels ‘stupid leprosy hit person’ or ‘not only animal, he is also a devil.’ Some of such labeling expressions could be due to the characteristics the labeled person is showing on social media (see earned reputation). In what we call venomous group attack, the troll insults not only the individuals who cause the irritation but also the entire (ethnic) group they belong to. In such cases, the troll disparages either by directly calling the specific ethnic group or its symbolic ties such as tradition/cultural items, language, community figure, as seen in the following comments given in response to an individual’s action: ‘Ethiopia has been bled by **these termites!** The problem with us is considering **them** like human beings’ (bolded to emphasize). In this excerpt, even if the irritation came from a single individual, the use of the word ‘these’ and ‘them’ shows the troll went beyond attacking the irritator per se and included the group the irritator is a member. Table 14 show instances of comments that show venomous group attacks:

Table 14 Sample Venomous Group Attack

S/N	Excerpts
1.	Alas Agame; you consider yourself as human and say you helped others to be human
2.	It is true that these people think like animals
3.	It is Qeero and its father-OLF that are known for killing humans and robbing Banks! Isn’t it???
4.	Son of Dogs! You robbed the country!

The second types of trolls identified in this study are *parrots* who characteristically repeat the same labeling used by outgroups to insult back their counter group. If one’s group is insulted with an offending word, parrots use that same offending remark back to that individual or his/her group. For example, an insult ‘lousy’ and ‘beggars’ directed to a group, its members have seen replying ‘it is you who is lousy’, and ‘you are sons of beggars’. In the same fashion as venoms attack either individuals or groups, parrots also do the same. In essence, parrots are characteristically engaged in tit-for-tat who just retaliates with same type and amount of weapon. In behaving destructively trolls can be opinion leaders or followers/echoers, consistent and random or inconsistent and sporadic.

The reflection made on the overall journey of data collection, analysis, and memoing shows negative comments can be given to four types of main posts which can be depicted in implicit-explicit and positive-negative dimensions, as seen in the diagram below:



The first one is bold and blatant hate speech: names ethnic groups, and use direct derogatory words. This is more of incident based hate speech. The user steals agenda by using available episodes to derogate its targets. Second, textual hate speech: speeches disguised within historical narrations-oral or written. The source seems avoiding responsibility for the speech. Third, apparently positive message but implicitly has elements that can drag individuals to hate. For example, the speech appreciate a group in a way its relevant outgroups could feel implicitly offended. The appreciated group shows arrogance in a way that others react negatively. The source is usually individuals who want to indirectly attack outgroups. Fourth is positive message such as building schools or railways somewhere in the country. But individuals are seen reacting negatively to such posts, presumably seeing conspiracy or based on their prior attitude.

4.2.2 Pace-making

While trolls mainly disparage others, other social media users are characterized by contributing *additional* conflict supportive remarks or ‘sensationalizing’ the topic under discussion, or by encouraging the hate monger who posted the hate speech content or the troll. I named such roles played by social media users as ‘Pace-making’ to imply that they add fuel to an already posted hate speech, to inflame further verbal fight. It appears that pace-makers are strategist and reserved than trolls are. For example, the data shows trolls are blatant in attacking targets while pace-makers direct their message to the trolls whom they support. Pace-makers can play one of the following roles: encouraging offenders or trolls; discouraging peacemakers; inviting audience to the hate circle such as by adding further toxic information, or obstinately blocking productive discussions.

Under the heading of pace-making, we have identified the following three subthemes. The first and most noticeable one is motivators, who are characterized by blessing, appreciating and encouraging trolls or the source of the hate speech. In encouraging the troll, motivators mainly follow two roots: appreciating the troll or appreciating the hate remark (or both sometimes). In appreciating the source, the pacemaker praises the troll as intellectual (e.g., ‘you are a moving library’), a hero and a patriot, a man of truth and/or thanks or blesses him/her (‘you made my heart happy’, ‘May God bless you’), for the toxic remark made by the troll. In appreciating the disparaging remark, the pace-makers praise the content of the hate-filled remark as truth, correct and something to be appreciated, as seen in the following social media excerpt given to an apparently hate-filled content: ‘I’m listening to you forgetting it's time for bed’; ‘this is the complete truth and nothing but the truth’ ‘you spoke what was on my mind. The following, a screenshot comment, from Facebook illustrates this role of social media users.



The screenshot above contains two comments showing appreciating a social media user who shared a video in which he is caught directly insulting the Oromos while expressing his anger against the killing of innocents by presumably Oromia Regional Police, at Weybela St. Marry Church, on January 19, 2022. The video received 8.8 k reactions, 4k comments, and 111 k views by the time the video is saved. While the source is heard repeatedly insulting the entire Oromo, many of the social media users who reacted to the video were appreciating the troll.

In online transmission, motivators show their presence by greeting the troll (to tell they are attending the live transmission), receiving orders from the troll, requesting actions (such as blocking those who disturb the online transmission) or asking for more similar content (e.g., ‘you have to write a book regarding their barbaric nature’, ‘you must expose the hidden evils more and more...please’, ‘I watched it twice, and I still need more of it’), or in parroting the exact words of the troll to show how they are impressed with the remark. For example, borrowing the words from a troll in a hate video, a motivator is seen commenting that same

phrase ‘ወንዝ ለወንዝ መቀደስ ያረፈ ሰይጣን መቀስቀስ; ካልጠፋ ቡኸር ነብር አደን!... Wow! Well said)’. This Amharic quote is directly borrowed from the troll in a hate filled video who disparages its target. It should be noted that the above comments are in response to hate speech.

The *Inflamer* is the second subtheme identified under the Pacemaker and is characterized by adding fuel to an ongoing hate-filled discourse, by providing additional toxic topics or by eliciting negative emotions, or reminding the troll and the audience of supporting accusatory claims as seen in the following quotes: ‘አጎርሳ በስሟ ብቻ ማወቅ ትችላልህ ሚዲያሳ ሲል ነገር አለቀ አገራችን የተማረ ሰው ጠፍቶ ነው ካሜሪካን የተጠራችው ለምን ይመስልህል ለዚህ ነው አትጃጃል ወገኒ ያለቀ ጉዳይ ነው’), which can be translated as: ‘note this, her name speaks volume! Why is that she was called from USA while there are many educated individuals inland! Don’t be fool as the plan is almost over’. While receiving additional anger inducing information, the already aroused audiences are more likely to validate the previous hateful idea and hence will be more irritated or angered. In addition, such expressions attract both ingroups and outgroups towards the hate speech circle. Excerpts such as ‘The late PM Meles was their master who once said "Aksum is nothing for Wolaitas"; 'Wollega **too** belongs to Amhara,' and 'we **will never forget** their saying 'it is we who emancipated the people (of Ethiopia) from living the life of apes on trees ’ show this role.

These three consecutive comments above, which are remainders of additional toxic information, are given just to add fuel to an already heated tug-of-war like verbal fight (i.e. hate speech) on the social media, and support the main source of the hate speech. While inflamers add new topics for discussion, they also direct audience’s attention to the toxic information being discussed, 'listen carefully; they are calling themselves 'federalists' while labeling you as terrorist.'

The second way the inflamer fuels the discourse is through emotional venting on the platform (i.e., awfulizing), in a way that is contagious enough to influence others, such as expressing emotions that show how one is angry, bored of the situation, fatigued, hopeless as seen in the following comment, 'ohhh the evil done on us is limitless.' While this type of emotional releasing appear normal and responsibility of members’, they are more likely to contribute to the strengthening of sense of victimhood and its associates.

The third subtheme we identified under pacemaker is the *stubborn-obstinate*, who react rigidly and are combatants to the point of inviting verbal fights, as seen in the following four excerpts:

Table 15 Sample Stubborn-Obstinate Reactions

S/N	Quotes
1.	Keep your mouth shut! You bring nothing, coward;
2.	'Whatever you say, you can't change anything'
3.	'Wrong analogy! 'Keep this analysis in your wallet.
4.	Yes, Neftegnas backbone is broken; the land has its owners. If you don't like to see Oromos, you better go to a place you like

The stubborn-obstinate subtypes also include those who directly reject other social media users' claim without explaining why.

4.2.3 Peace-Making

Peacemaking is the third type of role played by social media users in reacting to hate speech contents. Peacemakers are distinguished by their constructive contributions to cooling the heated conversation and hence to the peacebuilding process. In the *Coolers-Pacifiers subtheme*, the user directs attention to calm down bad feelings and transform hate-filled talk into neutral or uplifting situations. They peacefully communicate to the source of the hate speech, or to the audience on the platform. They also provide the audience advice on how to control their emotions, not to be gullible, not to hate or insult people, not to spread the hate post, to pray to God for peace, to have fruitful conversations, etc.

For the source of hate speech, they request them to have constructive conversations. The cooler-pacifiers also commands the attention of the troll to one's conscience and to reassess the destructive online behavior, or make an appeal to their mortality, ask not to incite hatred among people, or discourage the hateful message of the trolls without offending them. The following five quotes, which are comments given by different SMU to different hate speech posts, highlight this theme:

Table 16 Sample Peacemaking Reactions

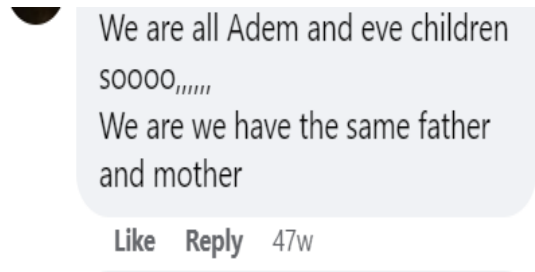
S/N	Sample excerpt for peacemakers
1.	Please stop spreading hate among people
2.	my brother, Ethiopia needs our prayer
3.	have we forgotten that we are mortal beings?
4.	we don't listen to you as Tigray people are our brothers

The screenshots below, which are comments given to hate speech posts, also show peacemaking role of SMUs.

ልብ ያለው ልብ ይበል!!

በሚገርም ሁኔታ እሮሞው በጣም የተጋባውና የተዋለደው ከእማራው ጋር ነው። እማራውም እጅግ በጣም የተጋባውና የተዋለደው ከእሮሞው ጋር ነው። ይሁንና ከሁለቱ ህዝቦች የተገኙ ምሁራን እና አክቲቪስቶች ከሐቅ ይልቅ ለስሜት እየተሸነፉ የሁለቱንም ህዝቦች ጥቅም በሚጎዱ ተግባራት ላይ መሰማራትን ስራቸው አድርገውታል። በተለይ ደግሞ ዛሬ የተናገሩትን ነገ በእያፈረሱት የጭራቸው ዳካ አልያዝ ብሎናል።

ወገን ትሰማኛለህ?? እይታህን አጥብብህ ግለሰቦች ለሚናገሩት ነገር ትኩረት እየሰጠህ በረባ ባልረባው አትናቆር። አስተሳሰብህን ሰፊ አድርገህ ስትራቴጂካዊ በሆኑ ጉዳዮች ላይ አተኩር!!



The excerpt in the left above can be translated as,

Surprisingly, marriage rates between ethnic Oromos and ethnic Amharas is higher than other ethnic groups. However, the ruling classes on both sides choose to overlook this fact and do things that are detrimental to both groups. They contradict what they say in public, which greatly confuses us. So, please refrain from listening to such people's [damaging] speeches and engaging in contentious debate in vain. Try to broaden your perspective only on crucial strategic concerns.

The other subtype of peace-maker is the *sympathizer* who characteristically feels pity for wrongdoing of the haters and the gullibility of social media users, expresses his/her worries about the consequence of being credulous to a particular hate speech post, and provides advice or helpful guidance. The following three excerpts clearly show the sympathizing role of social media users: ‘the government should calm down the people; the intention of the offenders is to create such chaos in the country; so, the youth must calm till the killer is caught; rest in peace brother’, ‘May God give you the heart to love others for your future life’, and ‘it is unfair to attribute every problem to PM Abiy Ahmed’. The sympathizers are also seen praying for both haters and victims, as seen in these comments: ‘please don’t touch the poor Ethiopian’, God forbid!, ‘Forgive their transgression as they don’t know what they are doing.’

Occasionally, I came across with unrelated but "good" messages, such a non-offending joke, details about oneself or one's own social media page, advertisements for scholarships and

businesses, and football news within a string of derogatory remarks in a way that breaks up the toxic conversation going on among social media users! Because such qualitatively distinct engagement of social media users has the ability to divert the attention of others from toxic rhetoric and disparaging remarks and command their attention to presumably funny or neutral contents, I have designated such roles as *detour finders*.

4.2.4 Informing

This role of social media users is characterized by the tendency to appear as knowledgeable or an expert on a subject under discussion and with a perceived mandate to update audiences about important social issues. This role primarily includes activists masquerading as journalist and genuine informed citizen or posing as such who appear as representatives of their respective ethnic group. The *interpreter* is the first subtheme identified under this role, who appear to be insightful and hence engages in interpreting what one calls ‘hidden’ or ulterior motives of the relevant outgroups (i.e., conspiracy reader) and inform their interpretation to their respective audience-ingroups. By doing this, interpreters could contribute positively by informing their audience what they might not be aware of otherwise so that they could exercise caution against falling prey to disinformation and hate mongers. The following excerpts illustrate this function, "This is a strategy he use to lend Tigrayans to Hatred; you better support the poor people", ‘I know you are voice of stupid Junta who pretend to be Amhara’, ‘Abe please look at this video link carefully, I suspect this man is not genuine/healthy’, and ‘you don’t represent ethnic Amhara, you are a paid hate monger striving to cause conflict between the two groups.’

The interpreter could also play harmful roles by (mis) interpreting others' messages or deeds in a way to annoy targets or damage the social fabric of the society, as evidenced by the following quote: "They must be kidding if they tell you that they want to help you in the Wolqite¹⁰ case while brutally killing you in Shashemene town and Wollega, and Arsi¹¹." In addition, in the screenshot below, a prominent figure is interpreting the words of another prominent figure from a different ethnic group and thereby ostensibly distorting the source's intent.

¹⁰ Disputed land that Amhara and Tigrayan ethnic groups both claimed ownership of.

¹¹ Administrative areas in Oromia region of Ethiopia

ገጽ 10
የአዲስ አበባ ስልጣን ስልጣን ስልጣን ስልጣን
አንድን ስሙት!
"እነዚህ ሰዎች (ትግራይ) እና
እነርሱን የመሰሉ
ሰዎች (አሮሞዎችና ሌሎች
እነርሱን መሰል ለማንነታቸው
የቆሙ ሰዎች) በኢትዮጵያ ምድር
እንዳይፈጠሩ ማጥፋት እና ስለ
እነርሱ የተፃፉ የታርክ
መዛን-ብቶችም ጭምር ማጥፋት
ብቻ ሳይሆን ከሰው ሂሳብም
አስተሳሰባቸው ጭምር መጥፋት
አለበት። ስለእነርሱ ማወቅ
የሚፈልግ ካለ፣ ተገሎው አልቀው
ከመራት ብቻ ተቆጥሮ አፅማቸው
ብቻ የሚገኝ መሆኑ ማረጋገጥ
አለብን። "

The above Amharic text could be translated as:

‘Please listen to it again if you don't understand what Mr [REDACTED] stated. He said 'We have to eliminate the Tigrayans, Oromos, and other ethnic groups who are fighting for their ethnic identity; we must destroy them not only from history and historical documents but also from our minds. Their thoughts must be eliminated from our minds.’

In this sense, the interpreter engages in "disambiguation" to make a point ‘clear’ while one is actually misrepresenting the source's original intent, and hence fueling hatred. This specific role is also considered as *disinforming* audiences with clear intent to deceive audience or cause intergroup hatred. It appears that the speeches of high-profile names are prone to such kind of (mis) interpretation. In addition, in its extreme form the interpreter's role becomes *prophesying* who foretells what is going to happen sooner or later. The prophecy maker, for example, tells the audience that the out-group is preparing to do something against in-groups and urges members to be ready for the ‘imminent danger’. In doing so, the interpreter tries to instill hopes and fear in the minds of the audience and urges them to take some kind of actions, contributing to a phenomenon called ‘accusation in a mirror’.

The second role under ‘informing’ is *inquiring* wherein the social media user challenges others to examine their position or the truth; seek more information, facts, or evidence. The inquirer is seen to follow a variety of patterns such as urging the others to read more so as to become more aware of the reality to come to their moral compass than messing the others with toxic messages. Yet, this role is not like advising; rather it is like one-upmanship. In a slightly different form, the ‘informer also requests others to undergo self-evaluation of their assumptions, idea, or judgment, as seen in the following excerpts: ‘what will be the fate of Tigray if you destroy Ethiopia? Please think...’, ‘is that taking care for Amhara? you are just paving ways for further massacre against Amhara’, ‘more than the source your role in

disseminating the speech is more harmful, please think of it’, and ‘what is the purpose of this post?’

4.2.5 Guarding

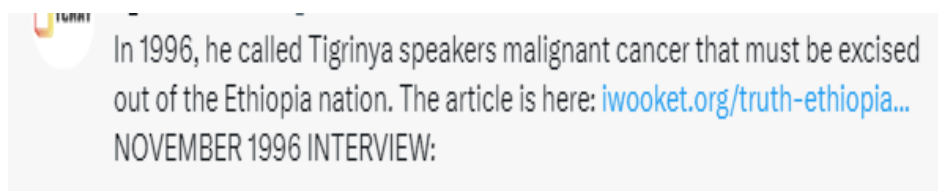
The tendency to care for and watch over the in-group is what distinguishes typical roles played by social media users, *the Guardians*. This role includes taking proper action that can ensure the physical, social, and psychological wellbeing of a group. In *guidance and direction giver*, for example, the guardian not only provides information but also guides ingroups what to do or not to do for the sake of their group’s wellbeing. In playing this role, one can ask members to share (or not to share), block, be wary of others on the platform, or take other coordinated/shared action to maximize group’s benefit, as seen in the following quote, ‘please do not share such kinds of posts, they are harmful to us’.

Another subtheme under safeguarding is *documenting*, who either records or insists on others documenting and sharing hate speeches directed against their group, so that the content reaches relevant partner/s/ who can take action, such as speaking out for them or filing legal action, sooner or later. The excerpts below shows this documentation role played by social media users:



The excerpt on the left reads ‘don’t comment on it but record and document it for history’; while the Tweet on the right is a response given to a video that portrayed TPLF as cancers (see page 85). The common issue for both comments is recording a the speech targeting their group so that others, including the next generation, could get and use it as evidence.

Another illustration for documenting and sharing hate speech is seen in the following Tweet:



Another subtheme found in this category is the image-builder who ‘defends’ the moral uprightness of in-groups, feels proud of in-group possessions, claims to always be at the

group's side; and assures audience that the future will be better for them. One also appears as defending the right and wellbeing of the in-group as seen in this comment "justice for our people."

While the main roles played by social media users in reacting to hate speech have been discussed above, the reflection I made (memo writing) on the overall journey of the data collection and analysis has given me the impression that two inclinations of users do exist: bystanders and reactors. Bystanders are those users who have the inclination to bypass offensive comments or posts. For example, the overall examination on data collection shows the presence of hate speech posts with some viewers but which cease to get any comment or reaction. It is interesting to examine why some hate speech posts immediately attract more comments, while equivalent others fail to get few or no comment while staying for longer period of time on platforms. Nonetheless, the observation shows there are bystanders who choose not to respond to hate speech posts for whatever reason they have. Bypassing is characteristically the modal behavior of bystanders on social media who cease to provide any form of reaction to hate speech posts against their own group or others. They opt to ignore hate speech posts and scroll down- no like, no share, no comment.

After describing the roles with their attributes, due to memoing and extended field visit, further exploration was extended to investigate how such roles contribute to the hate speech or to sustain toxic dialogue. This is seen in the number of negative reactions (replies) they received. For this, sample hate speech posts were selected and their comment threads are examined to identify the number of negative reactions each first role obtained. That means, from each main post the first comment representing each role was assessed after the data collection. Table 17 shows the summary of negative replies each first role obtained:

Table 17 Negative Replies Each First Role Obtained

Users' reactions	Main Hate Speech Post								Total
	P1	P2	P3	P4	P5	P6	P7	P8	
Trolling	3	-	5	3	1	1	2	3	18
Pacemaking	1	-	2	1	1	-	2	1	8
Peacemaking	-	1	-	-	1	1	-	-	3
Guarding	-	1	2	-	-	1	1	-	5
Informing	1	2	-	2	1	-	-	-	6

P= the main hate speech post from which comments are extracted

Table 17 shows peacemakers are not immune to receiving offensive reactions. Nonetheless, it is apparent that trolling has the most magnetic force to attract more toxic responses. It should be noted that the days each main hate speech post affect the amount of reactions they receive. That is to mean, as the hate speech post stays on the platform for longer, the more reactions and comment threads it will attract, due to their duration.

In Figure 11 below, the hypothesized link among the types of roles of users and occurrence and spread of hate speech is presented.

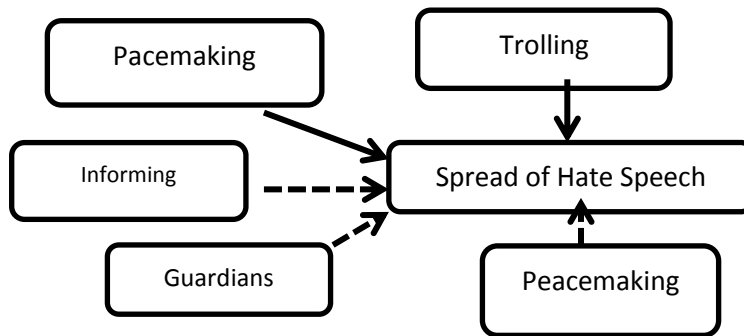


Figure 11 Proposed Links between the Roles and Occurrence of Hate Speech

As presented in Figure 11, the solid arrow shows a stronger causal link as the reactions have power to attract negative replies. The link indicated in the broken arrow implies there is a possibility that the roles invite toxic replies, though a weaker link is expected.

Section 4.3 How do Ethnic-based Hate Speeches Occur on Social media? (mechanisms involved in the occurrence of hate speeches)

With regard to addressing this last research question, both the dataset collected as well as the memo developed was used properly. It has to be noted that the lesson learnt in addressing the previous two research questions has guided the focus as well as the analysis pertaining to addressing this question. In this manner, some of the lessons learnt while addressing the previous questions are used both in designing this research question as well as in shaping my focus on a few salient issues that can inform mechanisms for hate speech occurrence. A major analytic tool employed to address this research question is asking from where particular hate speech content emerges.

The analysis shows three core categories, each with their own subcategories, which both describes and explains how hate speech occurs on the social media: earned reputation, poaching and trolling, and users’ naivety. Below, a nonlinear conceptual model of hate speech occurrence is presented followed by its description.

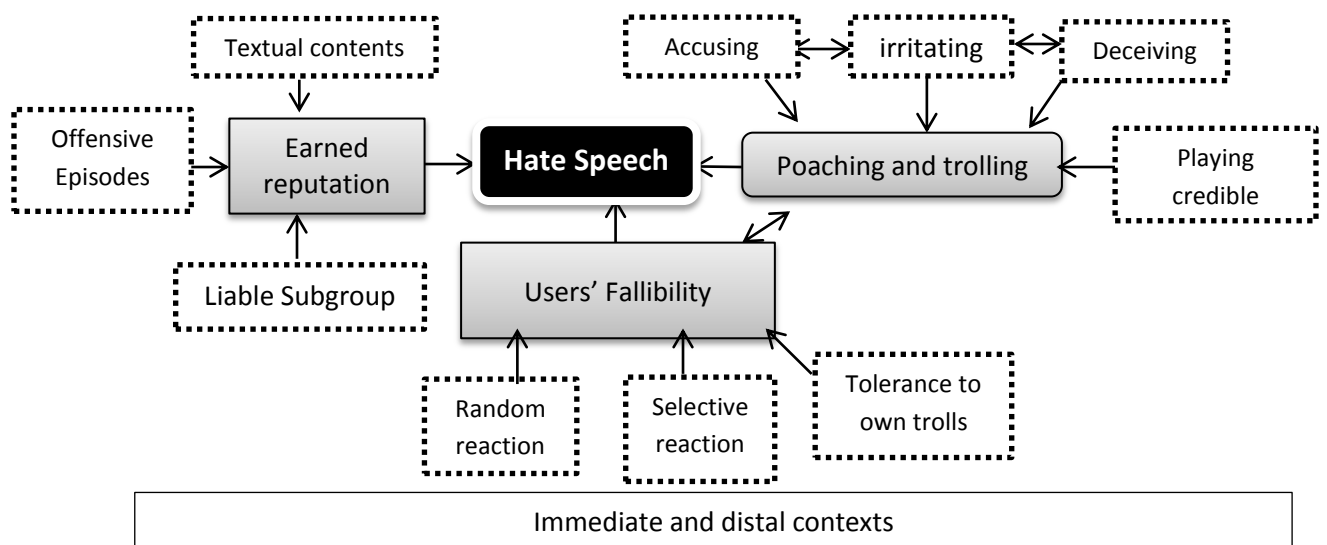


Figure 12 Proposed Conceptual Model of Hate Speech Occurrence

As the above model shows, the first core category is named ‘earned reputation’ which has three subcategories, ‘offensive episodes’, ‘textual hate’, and ‘liable subgroup’. The second conceptual category generated is ‘poaching and trolling’ with four sub-categories: ‘accusing’, ‘irritating’, ‘deceiving’, and ‘playing credible.’ The third main category is ‘users’ fallibility’ which has four subcategories: ‘random reaction’, ‘selective reaction’, and ‘tolerance to own trolls.’ The details of each core category and subcategories are presented below.

4.3.1 Earned Reputation: you deserve the negative label!

The data analysis shows ethnic groups are insulted because of the actual or perceived misdeeds, unappealing status of their members (including their elites), subgroups negatively labeled or hated by the government and relevant outgroups, as well as negative portrayal of them in various ‘historical texts.’ The mechanism regarding the occurrence of hate speech is, therefore members wrong action or misfortunes, hated subgroups and elites, as well as historical narrations in texts like books expose a group to hate speech. This hypothesized mechanism is referred in this dissertation as an ‘earned reputation.’ In this analysis, membership considers not only the present generation of an ethnic group but also the past generations (ancestors), and hence covers a wide range of time frame. Earned reputation as a springboard for hate speech assumes that there are reasons or justifications for negatively labeling a group: members’ heinous action or misfortunes, labeled or criminalized subgroups, and texts authored by elites that portray groups as criminal or enemies. The subcategories of earned reputation are presented below:

4.3.1.1 Offensive Episodes

The first subcategory of earned reputation is ‘*offensive episodes*’ which implies that ethnic groups are insulted, degraded, or described as enemies and threats because of the alleged or real misdeeds of their members including criminal and heinous acts, offending public speech (e.g., hate speech against others), and misfortunes (e.g. members seen as beggars). Offensive episode implies that ethnic groups become the target of hate speech, because of their own wrongdoers at present or in the past (irrespective of the truthfulness of the claim, though). For example, the analysis showed that evicting citizens from their residential areas, extrajudicial killings, and identity-base barbaric and vandalism acts which occurred in various corners of the country have lent ethnic groups (perpetrators are assumed to be a member) to hate speech. That is to mean, if certain identity based killing occurred in a particular place, the ethnic group dominating (‘owing’) that regional administration could be labeled ‘killers’ or ‘blood suckers’. In the previous presentation on types of hate speech, for example, the ‘worse than demon’ and ‘less than animal’ expressions emanated from actual heinous deeds committed by members of the group labeled as such.

These expressions not merely target the actual offenders but the ethnic group they are a part and are equivalent to expressions used by many Ethiopians to refer to parents of a rude child, ‘ሰድ አዳግ’ in Amharic language and ‘guddis-badi’ in Afan oromo language. In these expressions parents are insulted or at least implicated in the insult because of the rude

behavior of their child. In the same token, therefore, ethnic groups are insulted because of the misdeeds of their members. Taken from the dataset, the following three excerpts illustrate offensive episodes sub-category as a mechanism for hate speech occurrence.

Table 18 Instances of Offensive Episodes

S/n	Excerpt
1.	Crucifying human beings upside down and stoning to death; dragging dead bodies on the ground; burning churches and sacred books have become the traits of [REDACTED]. All these things are not immoral for [REDACTED]
2.	You sent your killers and made us orphans; destroyed our religious sites raped our daughters/sisters; massacred our youth and destroyed our infrastructures...
3.	[REDACTED] are devilish groups who kill drivers transporting food aid to them. We don't need to befriend or be neighbor with them. Their history is full of blood.

Offensive episodes as a mechanism shows that occurrence of hate speech involve at least three roles: ‘image abuser’, hate monger, and a target (ethnic-group derogated). The image abuser exposes its group, while the hatemonger utilizes that as an opportunity and justification to offend its target. Image abusers in this sense are individuals who, because of their actions or deeds, expose their own group for hate speech. Nonetheless, such path from wrong doers to hate speech could be based on unverified misdeeds as well as unrepresentative episodes. The offensive episodes as a means for hate speech also include presence of undesirable social status or public appearance (e.g., begging, publicly doing foolish things). In the previous section, I have presented that devaluation types of hate speech against Amhara happened because of portrayal of bare-footed villagers in Amhara region (see page 94). In addition, the venomous group attack types of trolls also clearly show how the hate monger was insulting the entire ethnic group merely because of the alleged misdeeds of a police office (incident of Weybela Mariam).

The public speech of elites, if others are offended by it, could also exposes their group to hate speech. For example, the speech, ‘we have broken Neftegnas backbone’ is used by many as evidence to assume and accuse the group the elite is a member are the same. That means elites’ public speech is taken as representing the group they deemed to represent. This is better illustrated by the following excerpts: ‘ማንነታቸውንና አላማቸውን አምልጦት ከተናገረው ንግግር ትክክለኛ ማንነታቸውን ነው ያጋለጡት እነዚህን ሰዎችን አምረን መታገል ይኖርብናል ከዚህ በኋላ ማንም አያምናችሁም’, which is

translated as: **‘the slip of tongue in his speech leaked their true identity and intention. We must bitterly fight these people** (bold to emphasize).

In this sense, hate speech because of offensive episodes employs a cognitive error called ‘faulty generalization’ from individuals to their group, implies presence of *image abusers within*, espouses ‘you deserve the insult’ and ‘call a spade a spade’ motto. It is also likely that because the observable misdeeds are used as a justification, earned reputation serves as a license to derogate target groups. Some of the offensive episodes analyzed in this dissertation include:

- Killing of innocents by the unknown groups (in many corners of the country)
- Publicized Criminal Offenses by high-profile names
- Killings by armed groups, such as during the Conflict in the North (e.g., Mai kadra¹², and slaughtering ENDF and Tigrayans- irrespective of one-sided claims and accusations).
- The alleged rape, immoral actions, and related brutality reported during the conflict in the North (some could be accusations!)
- Displacement of Citizens
- Ruin of Shashemene Town

The offensive episodes as a source for hate speech imply that lack of law and order contributes to the rise of hate speech. Partly, nonetheless, what I can call gray areas in politics play roles in the failure to uphold law and order. For example, the excerpt below is a comment given to a social media post that describes the burning of Ethiopian flag (with no star at the middle) by presumably a government police/soldier. Nonetheless, the social media user considered the doer as belong to ethnic Oromo and given the following remark¹³:

It is this group who slaughter, eat the flush and drink the blood of Amharas who are Orthodox church believers, both in Oromia region and Amhara region. To your surprise, no external invaders no Mosoloni of Italy, nor Ziyadbare of Solmalia invaders disgraced Ethiopian Flag this way; But the Oromos did it blatantly! Oromos emotion and arrogance have become unmanageable.

¹² A town in northwestern Ethiopia, wherein around 600 civilians were reported to have been killed in an ethnically-driven *massacre, during the Conflict in the North*

¹³ <https://twitter.com/AltayeEthiopia/status/1709685734611865770>

4.3.1.2 Textual Hate Speech

The analysis shows hate speech could be sourced from documents or texts that depict forefathers/ancestors of target groups as committing crimes or other forms of violence against ingroups. Such texts or documents could also portray a group as being invaders, settlers, land-grabbers who exploited and assimilated ingroups in the past. As such, hate mongers can draw such toxic textual contents from memory lane and use to frame hate speech. This takes us to our second subcategory of earned reputation-textual contents as a source for hate speech, which mainly is based on conflict supportive historical documents and related narratives.

The fact that textual hate speech is sourced from narration is well illustrated in the use of the adjectives inherent, historical, and eternal while calling the target group ‘enemies’. The label ‘historical enemies’, for example, implies the presence of narration or written documents that shows the group referred was, is, and will be enemy to the ingroups. In essence, it is not individuals who insult ethnic groups, rather texts / documents and narratives. In my analysis, hence, textual hate speech occurs when a troll uses written [historical] documents or dominant narratives to accuse, degrade, threaten, and dehumanize its target group, as seen in the following excerpt ‘ምኒልክ ሚልዮን ኦሮሞ እንደጨፈጨፈ ፋሲዳ ደራሲ በዛን ሳኦት በአይኑ አይቶ የፃፈ ‘ከእንጦጦ እስከ ባሮ’ የሚል መጻፍ አለለህ’¹⁴, which is translated as ‘you can read a book by a Russian author entitle ‘from Entoto to Baro¹⁴’ and which says a lot about the massacre by Minilik-II against the Oromos’. This excerpt show that textual hate speech can be accompanied by name dropping to convince others as well as to avoid any responsibility for their hateful expressions. In addition, it also shows that textual hate speech can use alleged offensive episodes as a means. The two pictures below illustrate how texts are used as means for occurrence of hate speech:

¹⁴ Referring to a book From Entoto to the river Baro, by Alexander Bulatovich; first published in 1897



Above, the video footage shows a man who is accusing Minilik II for oppressing and exploiting Oromos and the southern people while he was undergoing the alleged invasion, in the name of forming the current Ethiopia. While this datum was collected, the video had already received 1.4 thousand reactions, 182 comments (many are insults and mutual accusation) and 40000 viewers, all showing how widespread this message was. The picture on the right, shows a cover page of a book which portrays a particular group (presumably Amharas) as historical shrewed or ‘arrogants’. groups. While referring toxic texts, there are discussion titles, ‘ደሴው ሲገለጥ’ which means ‘when the file is opened’. Here, the ‘file to be opened’ is those with accusatory contents that portray a group as perpetrator and another as a victim. It seems that some users have clues about where the source of hate speech lies, as seen in this excerpt: ቁስ አባ ባህሬ ከፃፈው ጥላቻ የተሻለ ነገር አይፃፍም’, which means ‘no one writes more hatred than those wrote by Aba Bahre.’ The demonization type of hate speech illustrated on page 87 is also evidence of using documents to offend targets.

These kinds of toxic narratives and/or textual hate speech, irrespective of their truthfulness, is led by elites or educated individuals (historians, politicians, and activists) and can contribute to mutual accusation as each part brings its own counter narratives and attracts both supporters and contenders, as seen in the following Tweets:

 Replying to [redacted]
They discuss how the Ethiopian empire is birthed through Minilik's genocidal southward expansion & made legible to a repertoire of pan-Africanist & black liberationist movements as a progressive project

What emerges is a picture of a state that is a sign & signifier of genocide

18 55 120

 [redacted] @Habetopia · Sep 6
Replying to [redacted]
When are they going to discuss the Oromo expansion and the dozens of tribes that it wiped out from existence, among many, many....many other things?
I'm eager to read that one.

1 10

In support of this, many of the hate speeches observed in our study are framed in the context of disagreement over history, ownership over land (e.g., Addis Ababa, Welqite, Raya), and nation building (i.e., aboriginal-settlers, invaders) and use documents or narratives. The toxicity of these kinds of hate-fueled discourses is so harmful that it could also be seen as part of the *malinformation* (which may be true but harmful) category of the *information pollution*. The channel they use in playing this role includes formal discussion such as invited by journalists (shared on YouTubes), public speech such as in a rally, posting texts on the social media, and writing books full of accusations and labeling.

Nonetheless, such toxic ideas are echoed by ordinary social media users to offend targets. Each 'smaller' and 'fragmented' hate speech comments gathered from the various social media users appears to be on the same page as the main toxic discussion made by the high profiles (textual hate speech producers). I, hence, labeled the textual hate speech from the original source as 'primary textual hate speech' and the echoed one 'secondary textual hate speech.' In textual- hate speech, sources have little reservation in devaluing others, because they either believe they are presenting 'factual history' or just mentioning documents written by others.

4.3.1.3 Liable Subgroup

The third subcategory of earned reputation core category is '*Liable subgroup*' and shows how certain subgroups are a royal road and sometimes a legal means to offend the ethnic group they are a part. This means if either the government or outgroups openly detest or hate a subgroup that claim to represent the entire group, it provides haters rooms to offend the entire

group. The labeler could attribute his hateful expression against the identified subgroup on ground of real or perceived wrongdoing.

From legal, moral, and cognitive vantage points expressing hatred towards an entire ethnic group is distressing and unacceptable, perhaps because it makes one accountable by law, is morally questionable, and causes cognitive dissonance. In this study, evidences show that individuals feel they have the ‘license’ to attack the sub-groups or elites, especially when they are politically labeled as terrorist, criminal, or their faults publicized. The subgroup could be armed, politically identifiable, or are a culture-elements which are preferred like a soft meat to be easily attacked by others. During the conflict in the North, for example, many supporters of the federal government claimed they hate TPLF leaders or their fighters but not the entire Tigrayans. ‘I don’t hate Tigrayns, I hate TPLF’ was the royal road to post or share hate speeches during the conflict. Hence, it was very common seeing dehumanizing expressions targeting TPLF, but only in a way that offends nearly the entire Tigrayans. By the same token, the observation showed some social media users bitterly insult Oromo Liberation Front (OLF-SHENE), Qeero, Junta, and Fano, while reiterating that they merely insult/hate these special groups not the entire Oromos, Tigrayan, and Amharas respectively. Nonetheless, the overall analysis shows that nearly an entire ethnic group could feel offended provided that ‘its part’ is attacked or insulted, as seen in the following excerpt:



The data shows that the emotional connection between a sub-group and the entire group it is a part is not as loose as the labelers assumed and that offending a subgroup could be taken as an insult to the entire group. There is evidence that shows a significant number of members of the in-group feeling offended and reacted accordingly when their cultural marker, a sub-group, and elites are under attack.

Even if the word *Junta* was primarily assumed to identify armed TPLF groups-mainly the leaders, nearly all Tigrayans are identified with this term, calling themselves ‘I am Junta’. An

illustration is a video_in which a Tigrayan elite living in Addis Ababa shared this post in Amharic ‘ጁንታ ማለት የትግራይ ተወላጅ ነው (Gebeyanu, 5 March 2022)’ which means ‘Junta means Tigrayans’ and I am a Junta’. Another equivalent expression that illustrates this is the following excerpt from an unknown social media user: ‘ፋኖ ማለት አማራ፡ አማራ ማለት ፋኖ ነው’, which means ‘Fano is Amhara and Amhara is Fano.’ In essence derogating Fano would mean derogating the Amhara. Such path of hate speech from subgroup to the entire group is also facilitated by name calling, such as offering a special name. For example, TPLF, OLA, and Fano are named Junta, Shene, and Jawsa by the ruling government. Such name revisions further facilitate disparaging the subgroups, yet merely to offend more than those who are official members of the subgroups.

The sub-group-group path analysis as seen in liable group probably shows hate speech has collateral damage than hitting only the subgroup implied in the speech. This reminds us of a proverb in Amharic, ከተልባ ጋር የተገኘህ ሰሊጥ አብረህ ተወቀጥ’ which means ‘Toss together the sesame that you find with the flax’. The proverb, in essence, supports our observation that the entire group could be attacked because of its subgroups or part. This path of transference of feeling of offense, nonetheless, could occur through symbolic associations such as when cultural markers (e.g., traditions such as Ireecha) and political terms (Oromuma, Neftigna/rifle bearer, Unitarian (to mean anti-federalist), Minilikawiyen (to mean supporters of Minilik) are attacked.

In this sense, attacking parts of the group have a ripple effect to transcend into group level hate speech and that ethnic groups have something that expose them to hate speech. Our observation also assumes that individuals should have a stronger attachment to their group and a part that is under attack, so that our liable subgroup argument becomes convincing. More importantly, the more the sensitive part of the ethnic group attacked, the more it gets intense reaction and engulfs more and more individuals in the hate speech circle. Individuals are more likely to lend their ethnic group to hate speech when they appear in what is called ‘self-appointment’, which is acting as if one is appointed to represent the entire ethnic group, as seen in using the subject ‘we’ instead of ‘I’.

In my analysis liable subgroups are summarized in the following diagram.

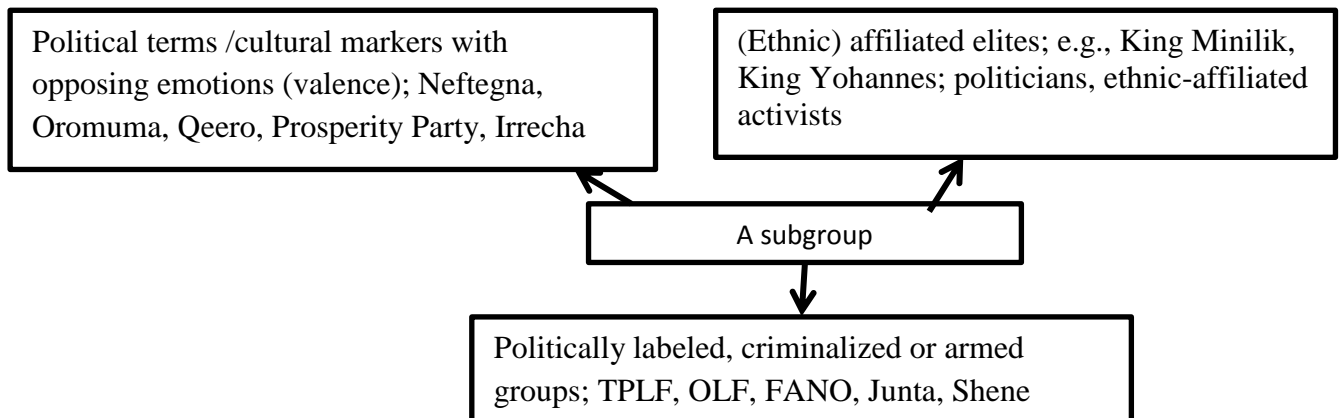


Figure 13 Subgroup as a means for Hate Speech

As figure 13 shows, the liable subgroup can be seen along three categories: political or cultural markers that are supposed to represent ethnic groups, ethnic affiliated elites both the living and dead, as well as ethnic affiliated political or armed groups. Because such entities attract hate speech I call them hate-bridges. Hate-Bridges are individuals, subgroups, and ethnic markers that an ethnic group is mainly insulted for. In this analysis, hate bridges are not inherently negative but hate mongers or relevant outgroup members use them as a gateway to offend the entire group. For example, insulting a legally criminalized subgroup (e.g., labeled as terrorist) is like having a license to insult it as well as the group it deemed to represent.

In conclusion, the conceptual territory of earned reputation is that ethnic based hate speech happens against a group because of the misdeeds or unappealing status or behavior of their members or subgroups both the living and the dead, as well as because of how they are represented in some texts.

4.3.2 *Poaching and Trolling*: intentions and techniques in hate speech

The second core category ‘Poaching and Trolling’ imply hate mongers on the social media who use various framing techniques to disparage their target while at the same time trap social media users to hate speech circle and hatred. These users act as the major agents for intensification of hate speech on social media.

In the previous section, I have presented trolls who offend others on the social media while providing comment to hate speech posts. Trolls are social media users who, sporadically or persistently behave negatively in insulting or derogating others but in the context of responding to hate speech content. In this section a unique set of trolls, named poachers, are

understood to exist. The data analysis shows some social media users are found to be lone-wolves and true haters who restlessly sensationalize social issues, consistently offend others, and hence play key roles not only in the occurrence but also in the intensification of hate speech. They appear to be motivated, persistent and more technical in offending others as well as in attracting more hate mongers to join them. I named this act among the social media users, *poaching* and the individuals who do it *as poachers*.

Characteristically, poachers sniff and exploit hot buttons, offensive *episodes*, hateful texts, divisive narrations, personal testimonials, dramatic stories, extraordinary occurrences, bizarre events, and slip of the tongue by members of the relevant group as evidence to substantiate their divisive claims and derogatory remarks. They are a lone-wolf in the sense that a lone poacher has the power to contaminate public spaces with hatred, just as a lone mosquito has the power to annoy a guy sleeping under a mosquito net. More importantly, they have well identified targets- specific ethnic groups they preferably hunt and attack on social media. In essence, all poachers are trolls but all trolls are not poachers.

These ‘serial offenders’ on the net appear to have a business to make whenever they are on the social media platform: fueling hatred. It is evident that poachers know how to frame hate speech so as to attract hundreds and thousands of social media users towards hate circle, which can be traced in the number of reactions they get and emotions evoked, while at the same time reaching towards their ultimate goal. Poachers’ characteristics can be described in the diagram below:

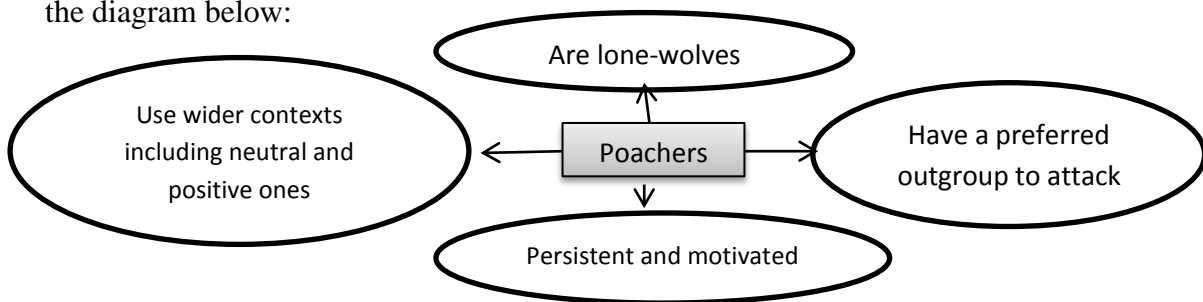
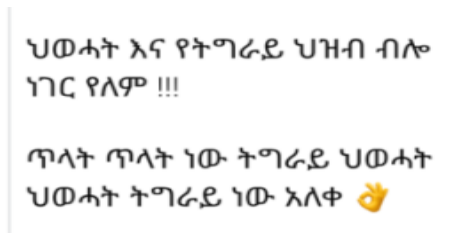


Figure 14 Poachers’ Characteristics

Poachers know sensitive parts of their own community/audience and have a wide range of techniques they employ to attract their audience. Poachers are known to use even positive contexts to disparage their targets. For example, poachers can use the Grand Ethiopian Renaissance Dam development, inauguration of a factory, or other apparently positive contexts to frame hatred and accusations. Regarding how they sensationalize messages and offend their targets, the followings are trapping techniques used by poachers:

4.3.2.1 *Accusation*

The first technique they employ is *accusation*, by which they portray their target as having ulterior motives (e.g., Oromos as separationists and Amharas as Unitarian), as criminals and oppressors at present and as inherent and historical enemies and threats to ingroups or third party. Accusatory contents in my analysis mainly accuse target groups for alleged land grabbing, political and economic marginalization, exploitation, and genocide. In the observation, ethnic Amharas, Tigrayans, and the Oromos accuse one another for alleged wrongdoing during the Minilik and Hailesilase era, TPLF-led era, and the prosperity party era respectively. Ethnic Oromos, for example, are displayed as ‘new oppressors’ (ተረኞች in Amharic) in the current Ethiopia. The word ‘new oppressors’, nonetheless, reminds us that there were oppressors in the past which are accused in other contexts. While employing accusation, poachers could engage in faulty generalization, as illustrated in the social media post below, which can be translated as ‘there is no difference between Tigray people and TPLF; they both are the same enemies. Period.’



It appears that accusatory messages have two sets of audience: ingroups and outgroups. Though evoking emotion, such kind of accusation is more likely to attract members from ingroups and outgroups as each want to defend their group (e.g., mutual accusation), and ultimately contributes to polarization that settling the disputes: no apology, no agreement! In many accusation instances we noted that the target also reacts with same weapon-accusing the accuser, sometimes leading to what is named accusation in a mirror (Marcus, 2012). In this study also it includes some element of this phenomenon, as exemplified by this excerpt: ‘አስታጥቁን አታስጠቁን’ which means ‘provide us with arms before they kill us’.

During the conflict in the North, some also portrayed Ethiopian National Defense Force, as Abiy’s soldier (by ethnic Tigrayan), ‘Neftegnas’ Soldier’ (by ethnic Oromos), ‘Oromuma Soldier’ or ‘Shimelis Abdisa¹⁵’s Soldier (by ethnic Amharas).

¹⁵ The current President of Oromia regional State, Ethiopia

As some authors argue, hate speech can be exhibited through symbolic attacks such as burning sacred materials and symbols the group values. In this sense, this picture is more likely to be considered not only as an offense but also a hate speech by relevant outgroups. Accusation, yet, could be based on generalized faults of individuals or could also be based on unverified claims as the above two picture may show. The following social media post along with its textual message (caption) illustrates accusations:



'It is believed that 90% of the newly appointed Finfinnee city officials are known anti Oromo people extremist Amhara ideologues who believe in the long-established genocidal ideology of the Neftagna System! Oromos everywhere must be aware of these worrisome developments, and fire up holistic liberation struggle for the self-determination & political sovereignty of the Oromo people, and territorial integrity of the Oromo nation!

<https://www.facebook.com/groups/2733367900232126/permalink/3094954094073503/>

While some accusatory statements are made to be in line with the dominant community narratives, others are none-falsifiable statements. We have also noted accusations used by poachers but they merely backfire by either exposing their weakness/ foolishness. In this sense, poachers are a shame to their groups and expose themselves and their group to hatred. Hence, while some accusations threaten outgroups, others merely backfire.

Poachers could also employ mal-information, which could be true but toxic. For example, poachers are seen resurfacing past actual misdeeds of outgroups or victimization of ingroups, presumably to accuse their target. In addition to this, during the conflict, poachers were posting accusatory messages which portray outgroups as secretly killing ingroups, as seen in the post below:

 5h · 🌐
 ዛሬ ላይ "የኢትዮጵያ መከላከያ ሰራዊት" የሚባል ተቋም
 የለም። ፈርሷል። ለዚህ ማሳያ በርካታ ነጥቦችን ማንሳት
 ይቻላል። ቫለቃ አህመድ አብረሃም, ቫለቃ ግርማ ፈይሳ,
 ገለጌ ሀብታሙ መገርሳ, ገለጌ አብዱሳ እዱኛ... የተባሉት
 የአቢይ አህመድ ጦር አዛዥ የአሮሞ ወታደሮች በማንነታቸው
 ተመርጠው፣ ፋኖ በተባለው ያማራ ነፍስ ገዳይ ቀማኞች
 (murderer mafias) ዴሴ ላይ መገደላቸው ትልቁ ማሳያ
 ነው።

 አቢይ አህመድ የሚባለውን ጨቅላ ሚንስተር
 የተፈናጠጠው የነ ዲያቆን ዳንኤል ኢትዮጵያ በዚህ ደረጃ
 ልጅ፤ የምትባሉ ነች።

 በዚህ ጉዳይ ላይ ይሄ [#FinfineeTimes](#) የጻፈው እኔ ካለኝ
 የውስጥ መረጃ ጋር የሚጣጣም ነው።

 ፋኖ ምን እስቦ ነው ...???

The post above accuses Fano of selectively killing ethnic Oromos who are members of the Ethiopian National defense force, during the conflict in the North. In essence, while accusing Fano and its supporters, it wanted to instill anger on its members

4.3.2.2 *Irritation*

The second technique poachers use to attract others is *irritation*. In irritating the target, the source seems wanting to elicit immediate anger response (reciprocal hate speech) from members of the relevant outgroup by employing degrading satirical messages or jokes that attack cultural markers (e.g., tradition, community figures), celebrating misfortunes of target groups (death of ethnic elites), and denying nationality or identity (e.g., ‘Oromos are not Ethiopians’, or ‘there is no Amhara in reality’) are typical techniques employed. It seems that ‘reciprocity principle’ operate to keep users in the circle of hate speech.

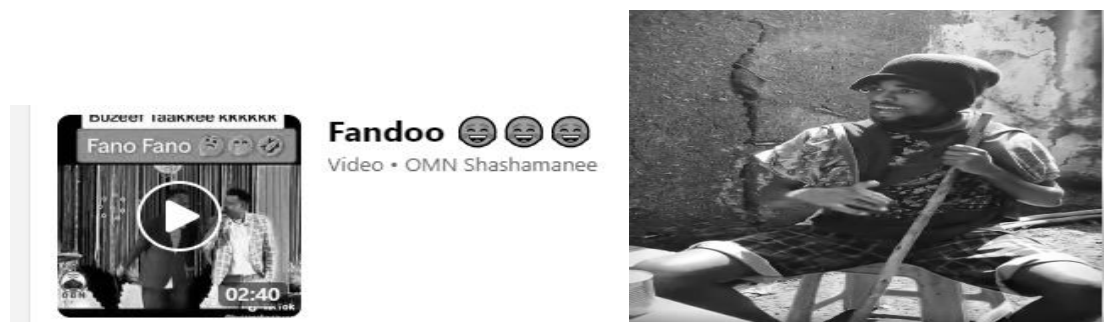
The following excerpt illustrates irritation as a technique ‘አሁን በየገጠሩ የሚኖረው አማራ ራስ ጃሽን ተራራ አካባቢ ከሚኖሩት ዝንጃሮዎች በምን ይሻላል ደግሞ የጥንት ስልጣኔ ትላለህ’, which could be translated as ‘Put aside your ancient civilization and tell me the distinction that exists between the rural Amhara people and the apes that inhabit the area around Ras Dashin Mountain.’ Such presumably irritating motives behind hate speech can also be exemplified by the post on page 94 which portrayed ethnic Amharas as aversive to wearing shoes.

Hate speech, framed to irritate targets is circulated at least partly to produce many likes and shares that will give rise to payment from platform companies. In this sense, teasing and mockery mainly using videos remind us the presence of hate speech which are meant ‘for sale’. This probably entails, beyond expressing attitudes, hate speech could also be used for making money, or perhaps for both purposes. In irritating others, the troll may not directly

offend others. Rather one can show up arrogance statements that enhance only one’s group (perhaps implicitly derogating outgroups). The following recent social media post illustrates this:



The above quote illustrates how poachers could behave arrogantly to irritate others. In addition to the texts, the emoji used in this post also shows the intention is to laugh at the target and cause irritation. Irritation could also be employed via derogating each other in ethnic jokes, including those which went viral presumably to make money out of it (hate speech for sale). The following video footages¹⁶ show how individuals ‘entertain their audience’ by derogating outgroups:



The two video footages are about music in which the trolls directly mention subpart of an ethnic name in disparaging manner, with the immediate impact of entertaining their audience. The video footage on the left, for example, is about two comedians who targeted Fano in their ethnic joke while their attendants are laughing loud. This video was broadcasted on a

¹⁶ <https://www.facebook.com/watch/?ref=saved&v=320271097081766>; <https://fb.watch/ndJRPEzA6v/>; <https://www.facebook.com/100002704583720/videos/657355822921013/>

4.3.2.4 Playing Credible

The fourth relevant technique we identified is *'playing credible'* which is characterized by appearing knowledgeable, truthful, freedom fighter, and right defender. While offending their targets or spread hateful messages, poachers typically take on the personas of thinkers, historians, journalists, activists, and fact-finders. In appearing as fact finder or a thinker, poachers appear well-informed about a pressing issue (i.e., they appear they have a secret code!) and use 'spade a spade' (ኅጭ ኅጫን) motto while narrating their divisive narrations. In addition, they also engage in 'conspiracy reading' and 'onion peeling' in such a way that actions by relevant outgroups are [mis]interpreted to fuel hatred or cause mistrust.

All elite poachers act in order to increase the credibility of their own demands and decrease the credibility of their enemy's demands. By same token, poachers also show arrogance with regard to the quality they want to show up publicly (i.e. upmanship). In appearing right defender, poachers divide an outgroup into two or more subgroup (e.g., Oromos into Shewa, Wollega, Arsic; Amharas into Gondar, Gojjam, and Qimant; and Tigrayan into Temben and Adwa) and show support to a subgroup while offending or attacking the others. In one of the videos analyzed in in this paper, an apparently Tigrayan man was observed portraying ethnic Amharas as killers of Qimant community, and it has received 1.1 thousand likes, 1.2 thousand comments and 19 thousand views the time the video was captured.

In some of their engagements, none-falsifiable statements are part and parcel of elites technique. In many instances, I observed poachers engage in self-appointing, as seen in using the subject 'we' instead of 'I' while narrating their views on the social media. It suggests that the troll assumes some level of authority or responsibility without necessarily receiving permission from others to do things on their behalf. Presumably, self-appointment helps trolls create the impression that they are in and for the community they appear to represent.

It has to be noted that most poachers are also victims on the social media they themselves and their ethnic group were frequently mocked at, belittled, or accused. This means for any hate speech post they share, members of the target group answer back with same weapon-hate speech!

4.3.3 Users' Fallibility

Apart from poaching and trolling, social media users are readily available to join the hate speech circle or behave in a way that contributes to the rise of hate speech. The analysis also shows poachers are well aware of the existence of certain groups of social media who must

behave on certain ways to pronounce the hate speech on platforms. This is more prevalent in the existence of audience segmentation as seen in directing message to specific groups. This takes us to our third main category, which we named *'Users Fallibility*. Users' fallibility/naivety: conceptually it encompasses gullibility as well as willingness of social media users to contribute to the circulation of hate speech. Users' naivety is evident from the very action or omission of ordinary social media users on media platforms. This core category has the following subcategories:

4.3.3.1 Random Reactions

Random reaction implies users' tendencies to nearly randomly like, share, or comment on hate speech post as well as every other types of post on the net. Random reactors are those who never want to miss reacting to posts; they have the tendency to like, share, or provide reaction to posts. Some random reactors react but with no clear intention with regard to supporting or disagreeing a hate speech post and hence may support both hate mongers and peacemakers on the social media. In addition, this category of social media users reacts to a post even before they think over the content, or with little cognitive investment. Random reactors are not behaving because of the negative attitude they have towards outgroups. Rather, they are more likely to behave out of mannerism. They are not critical customers on the social media who rush to give response before they understand the speaker in context! They contribute to the popularity of a hate speech post by the very reaction they provide-like, share, comment.

4.3.3.2 Selective Reactions

Our second subcategory under *users' fallibility/naivety* is the *selective reactions* in which social media users select poachers and contents that deserve their reaction. In essence, selective reactors are not interested to every hate speech post on the social media. Rather, for them, hate speech coming from certain individuals worth reacting. Therefore, this type of user's fallibility implies the act of reacting to only relevant post or poachers. That means, as poachers selectively attack a hated target, this user also reacts to those who deserve the reaction. The reaction they give is, nonetheless a motivation to the poacher; presumably it shows the post irritates their target. In so doing, they offend others and prompt them to answer back in a way that brings hate to themselves and their group, as seen in this quote: 'መሀይም ይህንን ፅሁፍ ከፃፈ ሰዉ በላይ የለም የኦሮሞን ማንነት ገና ታያለህ አንተ ደነዝ በመሬቱ ላይ ስለተሸከመህ ምላሹ ይህ ከሆነ'. This can be translated as 'no one is more illiterate than the one who wrote this.

you will see how strong is Oromo; you moron. Is this the response your provide to Oromo who allowed you to live on its land?’

This is more likely to keep all involved trapped in the hate speech saga. It appears that selective reactors are more concerned with who said it than what is said on the social media. This encourages them to react differentially to hate speech posts and individuals, a phenomenon we called *selective sensitivity*. In our analysis, *Selective Sensitivity* means some social media users differentially react to hate speech contents not only on the bases of what is said, but also who said it. For example, our dataset shows, when an apparently ethnic Amhara man said something negative towards ‘woyane’ (allusion to Tigray People Liberation Front), an apparently Tigraysn man responded, ‘woyane means a Tigrayan *especially when it is said by an Amhara man*’ (italics to emphasize; see the Tweet on page 126). Here, one can question whether this user would get irritated and respond the same way had it been to members of a different ethnic group. I would say probably No; because there was fierce political dissent between ethnic Tigrayans and ethnic Amharas (among the elites) during the conflict in the North, users were more likely to react differently. The attention given and the decision made regarding whether to respond to posts probably varies along who the source of the post is. *Selective Sensitivity* implies that individuals categorize outgroups as relevant and non-relevant outgroups (perhaps politically) and become highly sensitive to hate speech posts that come from members of the relevant outgroup. Relevant outgroups are more meaningful and remain at the center of their attention regarding a particular discourse or rhetoric. The relevant outgroups appear to be deserving the reaction because they are competing groups; while the later are those whose post including offensive one could be either tolerated or ignored- because they are less relevant or less sensitive. Selective sensitivity is also evident in sensationalizing contested contents if they are promoted or shared by elites than by ordinary social media users. Selective reactors are more likely to be confronted and responded back. This is more evident in what is termed as ‘lik-likun nigerew’, and ‘Asegebalet’ (in Amharic) which means ‘you nailed it’ and are responses that show the responder has got a bitter response from the poacher or troll.

4.3.3.3 Tolerance to Own Trolls

The third subcategory of readily available preys is *tolerance to own trolls* that implies that those who direct their attention and interest to their own poachers, easily believe in them and show appreciation and support to the post or to the poacher. These are tolerant and *gullible preys* who encourage their respective poachers and share and like their hate speech post. Such

preys are loyal to and prisoners of their masters, and demonstrate stronger and blind sense of identification with the poacher. A manifestation of such individuals is too much reliance on supportive sources.

As seen in the previous section on roles of users, motivating types are known for their encouragement and support given to trolls, as well as defending one from any external attack. In a related manner, members of a group also cease to criticize a hate monger within. This also shows users worry much on who is insulted by whom than the offense per se. Therefore, they are more likely to keep silent when their members are offending others. The evidence presented on page 80 on animalization, clearly shows this subcategory (members were even applauding the hate monger in the meeting).

Chapter Five

Discussions and Conclusions

5.1 Discussion

This study sought to investigate and clarify the different types and mechanisms involved in ethnic-based online hate speech in Ethiopia and explore how social media users respond to such hate speech posts online. In so doing, it also explored the contexts in which hate speeches are framed. The study analyzed ethnic-based hate speeches framed in various formats like text, audio, video, and images through a qualitative and inductive approach. In the analysis, pervasive and richer categories are focused so as to capture theoretically and practically significant findings. In addition to addressing the initial research questions, the study identified four interconnected contexts within which ethnic-based hate speech is framed: competition over history and resource, the conflict in the northern part of Ethiopia, low diversity and tolerance levels, challenges in upholding law and order—all intricately linked to ethnic identity.

In addressing the first research question, the study has come up with four major types of hate speech expressions: dehumanization, enemification, devaluation, and attacking and repelling—each with subcategories and a clear conceptual territory. Although not archetypes, the findings regarding the three forms of dehumanization expressions—animalization, objectification, and demonization—are consistent with previous studies and models of dehumanization (Haslam, 2006; Leyens et al., 2007; Bar-Tal et al., 1989; Tipler and Ruschel, 2014). Considering the level of hate speech collected and analyzed in this study, I would say the time period wherein the majority of the data was collected was characterized by the rise of dehumanizing expressions in Ethiopia.

As presented in the previous chapter, this study has identified various forms of dehumanizing expressions made by wide range of social media users, including elites and religious leaders, who portrayed their target as animals, herds, termites, daylight hyenas to the extent of using the labels as nicknames at least temporarily depending on actual events happening offline. Parallel to previous studies (Lakoff & Johnson, 1980; Li & Hills, 2021), it seems to me that such dehumanizing expressions metaphorically imply certain characteristics assigned to the group referred, such as stereotypes held against them.

Needless to say, beyond contributing to societal polarization and enemy making process, online dehumanizing expressions, under certain conditions, can bring actual offline violence. As evidenced by various studies, for example, portrayal of humans as animal leads to various

forms of identity based violence such as genocide, deportation, exclusion, and discriminations (Morrock, 2010; Waldron, 2012).

While ‘less than human’ (Kteily, 2016) and demonizing (Giner-Sorolla et al, 2012) expressions are known in the literature, the current study has found ‘less- than-animal’ and ‘worse than demon’ expressions used to describe the target ethnic groups or their parts, mainly in the context of the conflict in the North and actual offensive episodes happening offline. In these less-than-animal and worse-than-demon expressions, targets are assigned severe forms of undesirable traits that are even uncommon in the animal, beast, or demon world. Such expressions emanated from actual heinous acts members of a group committing ‘earned reputation/offensive episode’ mechanisms of hate speech. This finding shows how individuals (image abusers) expose their ethnic group to hate speech. Moreover, elites were seen in participating producing such kind of extreme types of dehumanizing expressions, implying how the expressions were dangerous in terms of having the potential to cause offline violence (Bensch, 2014). I strongly believe that the less-than-animal and worse-than-demon or beast expressions are unknown in the hate speech literature. In this sense, the current study contributes to advancing the available literature by adding additional layers to conceptualizing dehumanization.

Objectifying outgroups contributes to downgrading them mainly through their physical properties (Calogero, 2012) as well as exploiting (such as enslaving, using them as objects) and seeing them as lacking in agency and subjectivity. Objectified individuals are also more likely to be pitied or ‘cared for’ (denial of autonomy) or they must be eradicated like useless and/or dangerous items like weeds or diseases or waste products (Smith, 2011).

With regard to demonizing human groups, Barker argued that it is the most extreme form of the language of contention, which is the main means through which perpetrators justify and explain their killings and expulsions (2007). In support of this, Giner-Sorolla and colleagues (2012) argued that demonizing outgroups creates mechanisms that enable and justify violence and the hardening of emotions against the labeled group or its members. As hate speech in general and demonization in particular could lead to action, demonized targets, ironically, are likely to become victims while the labeler becomes a killer or ‘demon’ (Skitka, 2002). This shows victims of hate speeches are hurt by their labelers at least twice, first verbally and secondly with actions against them.

The problem with hate speech in general and dehumanization in particular is it removes the moral prohibitions against violence and genocide, thereby allowing them to happen. Due to

the use of survival concern to justify calls for attacking and calls made for emotional hardening noted in this study, perpetrators are more likely to withhold moral concern towards their targets. Instead, they could feel morally obliged to take actions against their targets, such as attacking or discriminating against them. In support of this, Singer (1981) and Pratto and Glasford (2008) contended that when ‘we’ and ‘they’ are capitalized during intergroup conflict, humans tend to treat ingroups with more moral concern and give more value to their lives than to outgroups. Echoing this, Reed and Aquino (2003) argued that the more identified a person is with a social group, the narrower would be his or her circle of moral concern for other groups.

Dehumanizing expressions are more problematic when used by elites during conflictual scenarios (Carment, 2007). Elites' use of dehumanizing metaphors can instill hostility in others. This is more powerful and consequential during perceived or real intergroup conflicts, such as the struggle for political power. For example, demonization is used as a tool of institutional power in Thailand and in Africa (Sripokangkul et al., 2020). Dehumanization, which is the fourth stage of the ‘10 stages of genocide’ by Stanton (1998), evokes feelings of contempt and disgust, removes the moral prohibition against violence, and hence enables genocide and systemic discrimination. A study by Hodson and colleagues also highlights the role of dehumanization and the emotion of disgust in explaining intergroup relations, emphasizing the human-animal divide and intergroup disgust (2014). It seems to me that dehumanization, and hence the removal of moral concern, explains why identity-based attacks and heinous violence were rampant during the data collection period in Ethiopia.

The current study also shows how ethnic groups are portrayed as enemies and threats. Enemification of outgroups is believed to be a language pattern of outgroup prejudice and can elicit confirmatory behavior from members of a subordinate group via a self-fulfilling prophecy (Li & Hills, 2021). In the current study, the enemification type of hate speech is more related to victimhood narrations supported by texts that portray target groups as enemies and threats. In essence, poachers' accusation technique mainly portrays the target as enemies and threats. This type of hate speech is more likely to be supported by texts, disguised within discussions of historical events, and hence espouses ‘it is not me who said this’, while offending a target group and use previous writers as scapegoat. Enemification of outgroups is found to be the major obstacle that sustain retractable conflicts, such as the one between Israel and Palestine (Maddrell, 2015; Rieber, 1991).

Descriptions of others as enemy are, therefore, problematic for peaceful coexistence, as hate can emanate from and cause fear (Maddrell, 2015). A relevant story arising from fear is the subtle infiltrator theory, which maintains that a person is afraid that his or her safety and/or the safety of one's group are threatened by infiltrators of some kind (Sternberg, 2005). Fear is, in large part, socially learned and fulfills two functions in the development and transmission of hate. The primary function fear serves is to create an aversion to, and even hatred towards, a target individual or group (Sternberg & Sternberg, 2008). The ultimate goal of the instigator is to recruit as many potential perpetrators as possible. A secondary goal of instigating fear is to prevent observers from thwarting or, in any way, interfering with the wicked goals of the instigators. Since targets are described as killers, land grabbers, and assimilators, enemification also evokes intergroup anger and the desire to take revenge. Intergroup anger, in turn, predicts support for more radical and unconstitutional forms of action (Livingstone et al., 2011). Perhaps knowing this tendency, conflict mongers employ techniques that they think can evoke fear in relation to the relevant outgroup. Parallel to previous studies (Bar-Tal, 2014) the current study, has understood elites and /or poachers do disseminate conflict-supporting narratives and conspiracies, which mainly brings benefits to ethnic leaders (Carment, 2007.) Because enemification implies portraying groups as killers of ancestors, invaders, assimilators, and land grabbers, I have the impression that it is used as a political tool in the context of identity politics in Ethiopia.

Devaluation type of hate speech is highly tied to contempt and disgust (Sternberg, 2003), which is a language pattern of outgroup prejudice (Li & Hills, 2021). This type of hate speech probably is the manifestation of the "commitment" element of Sternberg's Triangle of Hate (2003). Some of the devaluation expressions in the current study are related to teasing such as joking at ethnic groups (e.g., through music), perhaps with the intention to express a true attitude or just for the sake of irritating the targets. In essence, this kind of devaluation is related to the concept of casual racism, which is a form of subtle or indirect racism that typically involves making stereotypical comments or jokes about a particular social group. When such joke against a group is antisocial in intention, it may reinforce stereotypes about the group it targets, potentially loosening societal norms that discourage expressions of prejudice (Saucier et al., 2016). Ford and Ferguson (2004) noted that disparagement humor creates a normative climate of tolerance of discrimination, as well as variables that accentuate and attenuate its effects. In essence, offending humor could desensitize the offenders, while as the same time irritating target members. Racial humor is a tool that divides societies by

reinforcing and perpetuating negative stereotypes (Maio et al, 1997). Its enjoyment has been used as a measure of prejudice toward target groups (Crandall et al., 2002). The devaluation type of hate speech evokes disgust and contempt and is mostly seen in relation to the irritation techniques used by trolls and poachers identified in this study. It seems that devaluation is mainly employed by ordinary social media users than elites. It appears that devaluation has a strong relationship with subtle attitudes and negative stereotypes that are simmering undercurrent and is same as the first step in the Pyramid of hate (Anti-Defamation League, 2005), which can grow to the next action component of discrimination and violence.

Outgroup derogation can serve a public presentation function that allows for the enhancement of an insecure status within a desirable ingroup (Noel et al., 1995). In addition, members of devalued groups may internalize negative stereotypes, and being devalued based on group membership can cause emotional distress, anger, a negative attitude, and avoidance. It is also possible that the target group will fire back same labeling against its haters.

Attacking and repelling types of hate speech observed in this study, which is the action component of the Pyramid of hate (Anti-Defamation League, 2005), calls for urgent attention, as the path that links such calls with actual violence, including genocide, is stronger, especially in the context of intergroup conflict. This type of hate speech is similar to what Benesch referred to as 'Dangerous Speech', which she defined as ways of communication that encourage acts of violence and/or lead to toleration or even exercise of violence by certain groups (2014). Evidence shows that, during intergroup conflict, calls made to register individuals based on their ethnicity, to enter outgroups into concentration camps, and emotional hardening are royal roads to genocidal incidents such as those observed in Rwanda, Germany, and Cambodia (Brennan, 1995; Smith, 2011; Soral et al., 2018).

More importantly, the actual and direct calls for attacking outgroups on the grounds of survival issues were dangerous calls that would force audiences to take preemptive measures to avoid the perceived imminent danger. In addition, such calls also have the power to justify any actions taken by members. In this sense, calls made to take actions against a target on grounds of threatened survival issues are nearly genocide-level calls (Ilibagiza and Erwin, 2006; Stanton, 1996) and resemble accusations in a mirror (Marcus, 2012). When mediated by outgroup dehumanization and exposure to crises further worsens intergroup hostility and predicts a lack of willingness to help the target (Adler et al., 2021). Denial of help and thwarting inter-ethnic marriage are issues that are worth attention as they could challenge

diversity tolerance in Ethiopia. Nonetheless, this study has also documented a new version of the expression of hatred: thwarting any help coming from outgroups. While the literature says a lot about the fact that haters may withhold help from victims, this study has found that those who feel victimized express that they don't need any help from members of the alleged perpetrators. This is observed in the context of the conflict in the North.

In addition to ordinary social media users, community leaders such as politicians, activists, journalists, religious fathers, and academicians have involvement in the hate speech observed in this study. These elites have social power to convince the mass and strengthen stereotypes and prejudices. To make it worse, top government officials leading regional government are observed in expressing demonizing and related hate speeches against groups or subgroups, especially in the context of the Conflict in the North. Because political elites in the context of Ethiopia are more likely to be affiliated to a particular ethnic group, the hate speech from them is more likely to be assumed to represent the ethnic group they are a member. This in turn contributes to polarization through generalizing faults and bad-mouth (elite's hate speech).

It appears to me that some of the pejorative and derogatory labels used in the hate speech expressions such as Ga*la, herds, daylight hyenas, junta, neftegna, Oromuma, and settlers are the result of the toxic narratives by elites and create unsafe environment for certain members it targets as well as for the country. Apparently, the hateful expressions observed in this study are harmful to peaceful coexistence as it fuels and justifies violence, and harden emotions against outgroups (Giner-Sorolla et al., 2012; Smith, 2011).

I contend that, because social media is iterant and keeps circulating over long distances, such expressions are more likely to be repeated both online and offline, contributing to making them cognitively accessible. Because of the 'linguistic warfare' seen in dehumanizing, devaluing, and enemification groups, and with repeated use, these hate-filled expressions assume self-evident status (Lakoff & Johnson, 1980), it desensitizes labelers to victims' suffering and provides social support for violent actions against the victims (Brennan, 1995). Drawing from the socio-cognitive traditions in social psychology, for example, it is argued that the language people are exposed to and the language they employ, shape the way they think and construct reality (Cervone et al., 2020). Needless to say, parallel to cultivation theory (Gerbner, 1998), hate speech expressions seen in this study contributes to the societal polarization, as its whirl attracts more social media users towards its circle. More importantly,

victims could believe that the relevant outgroups do hate them, which in turn tempts them to behave in certain ways such as defending themselves with equivalent hate speech expressions (Cervone et al., 2020).

It appears to me that publicly available online hate speeches are more likely to be shared with intention. One can question why the source left the speech public, despite the option to edit or delete content from the platform. Leaving such damaging hateful expressions on media platforms publicly suggests both the willingness and intentionality of the source as well as a lack of inhibition and social constraints. The persistence of hateful expressions may also indicate deeper societal and political tensions that could potentially escalate to real-world violence given the right conditions (Allport 1954; Anti-Defamation League, 2005).

Even though members of three ethnic groups are implicated in this study and that haters mainly have preferable targets, pieces of evidence show that people who reject one outgroup tend to reject other outgroups at least in a different time and context (Bergh et al., 2016). Taking the idea that hate speech towards groups is symptomatic of societal stereotyping, and othering processes (Allport 1954), it is reasonable to make use of online hate speech to understand and mitigate offline conflicts. Offline triggering events are often followed by increases in types of online hate speech that bear seemingly little connection to the underlying event. This is more powerful when conflicts and negative sentiments already exist. For example, negative beliefs regarding an outgroup enhance negative evaluations of the conflict situation (Giebels, 2008). Therefore, in order to avert serious repercussions during conflictual scenarios, peacemaking such as through countering hate speech should be done during relatively peaceful scenarios.

In addition, the hate speech expressions identified in this study are framed within a highly fertile ground (e.g., the conflict in the north) that can transcend into offline conflict, including genocide. General strain theory, for example, posits that various forms of violence including crime and delinquency, are caused by negative emotion(s), including fear, disappointment, anger, and frustration (Agnew, 2006). These negative emotions can be elicited due to repeated hate speeches targeting one's group. Nonetheless, even if Ethiopians have faced ethnic-based offline conflicts in various corners of the country, I contend that certain protective factors have prevented the escalation of violence, suggesting the presence of buffering factors. This argument, yet, should be understood with the recognition that the hate speech expressions observed in this study are more likely to have contributed to the actual offline violence, such

as the internal displacement, the identity-based killings, and property destructions we observed in Ethiopia over the past few years. Indeed, theorists suggest that politicians may manipulate messaging to facilitate the "other" or an "us versus them" mentality as well as supply hate-creating stories (Haslam et al., 2020). In support of this, Caspersen (2008) highlights the presence of divisions within ethnic groups, indicating power struggles among elites rather than solely differing views on protecting collective interests. Perhaps parallel to the above two arguments, Gagliardone and colleagues (2014) argued that the online hate speech observed in Ethiopia didn't reflect the actual intergroup relations on the ground—the offline intergroup situation is better than the online hate speech. Equally important to note, nonetheless, is the time frame in which the data was collected- highly conflictual scenario in Ethiopia. Viewed from this perspective, one may question whether equivalent hate speech expressions could be found before and after the conflict in the North. I do believe that the two year deadly conflict in the northern part of Ethiopia has drawn what otherwise would be decent and peace-loving individuals into hate mongering ones.

Nonetheless, the hate speeches identified in this study requires attention and action to avert their polarizing effect as well as offline consequences. Goroka, for example, argued that target group may feel the hate speech against them reflects the thoughts and social attitudes of that group (2019), thereby contributing to group animosity and averting the continuum of destruction that awaits them otherwise.

I strongly believe that the four types of hate speeches generated in this study are likely to exist behind the various negative intergroup attitudes, prejudice, and negative stereotypes, as well as identity-based violence like exploitation, deportation, xenophobic attack, and genocide. Nonetheless, while the four types of hate speech expressions are supported by previous studies, none of them, as an independent study, has identified the four types as this study does. In addition, the less-than-animal and worse-than-demon expressions as well as thwarting help from members of outgroup (alleged perpetrators) are unique findings that can contribute to the available hate speech and intergroup hatred literature.

This study also came up with five major types of roles played by social media users in relation to reacting to hate speech posts: trolling, pace-making, peacemaking, informing, and safeguarding, as well as an invisible role-bystander. Although an individual may assume any of these roles in various contexts, some may frequently display a modal role across contexts,

perhaps due to personality traits and personal goals (Brogaard, 2020; Lampropoulos et al., 2022; Seidman, 2020; Katz et al., 1974).

Trolls are the main polluters of social media platforms and agents of hatred as seen in this study. Trolling and equivalent terms such as ‘hate mongering’, conflict mongering’ (Benesch, 2014), cyberbullying (Chavan & Shylaja, 2015), and online harassment and insulting (Cheng et al., 2017) have been reported in previous studies, generally to refer to internet users who disturb the media ecosystem by offending others (Coles & West, 2016). While trolls and equivalent terms in other studies imply individuals surfing the net and disturbing, insulting, or irritating others, the trolls identified in this study primarily imply users who derogate others while responding to the already posted hate speech by others.

Though trolling in this study is identified from the comments users provide to hate speech, the main hate speech post from which the comments are extracted and the four types of hate speech discussed earlier show this role of users. In addition, the overall observation and analysis (including memoing) showed that trolls can also be found while commenting on positive or constructive social media posts. This perhaps shows trolls may be more influenced by their preconceived notions or motive than by the actual posts there (Lampropoulos, et al., 2022).

Even though this study didn’t identify the reason for trolling, previous studies pointed out that trolling could be driven by prejudice, traits, fun, attention-seeking, or revenge (Shachaf & Hara, 2010). Hence, some trolls might be naive users who are unwittingly caught up in the vortex of hate speech, while others could be true haters. While Cheng and colleagues noted contexts can turn nearly every user into a troll (2017), the well-versed dangerous speech expert Benesch argues trolling increases in the context of high intergroup conflict, political instability, and economic strain (2014), which were visible in Ethiopia during the data collection for this study (European Institute of Peace, 2021; Muluken et al., 2021). In addition, Kinfe (2017) argued that the majority of social media users in Ethiopia are naïve and conflict mongers. The excerpt presented in the introduction section, on page 8, supports the idea that naïve and peaceful individuals can be dragged into being a troll in the context of the identity politics in Ethiopia. Nonetheless, studies show that with repeated exposure to online hate speech, naïve users (including naïve trolls) can be affected to see the world through what Gerbner (1998) calls a ‘televised viewpoint’, and hence will be desensitized, and gradually transformed into consistent haters (Cheng et al., 2017; Soral et al., 2018). In

this sense, the observed hate speeches and growing number of trolls in Ethiopia could be understood as the result of repeated exposure to hateful online and offline environment.

I assume that three interrelated reasons can be mentioned for this: First, as seen in parrots and consistent with social identity theory (Tajfel & Turner, 2004) and the self-fulfilling nature of hate speech (Article-19, 2015), trolls (including the main source of the hate speech post) are more likely to attract others to offend them (Cheng et al., 2017). The exchange of negative words ultimately feeds their cognition and contributes to the development of negative attitudes. This is more plausible in the context of ethnic-based hate speech, wherein identity is attacked and defense is automatically called for. It seems to me that, given the ethnic-based hate speech posts from which the comments are collected, social identity is more likely to be at the center of trolling in this study. Consistent with the idea that hate speech is self-fulfilling and self-reinforcing, therefore, ‘the reciprocity principle’ operates to engage in trolling, which is powerful in this context where identity politics is a driving force.

Second, consistent with self-perception theory (Bem, 1972) and the dissonance reduction of Festinger (1957), a behavior demonstrated once (e.g., insulting others for the first time) could lead to the development of an attitude supporting that behavior, which in turn could transform a naïve user into a troll. Third, trolls reinforce negative perceptions about their groups by badmouthing others, which can be used as evidence that their group is hateful (Lakoff & Johnson, 1980). Perhaps a fourth explanation is that users, through repeated exposure to hate-filled platforms, individuals learn the art and culture of offending, irritating, and devaluing their target. For example, users slowly learn the jargon, sensitive parts of the target group, and hot buttons used in framing the hateful expression.

It should also be noted that observing one's social identity (ethnicity or ethnic markers) being repeatedly insulted or attacked would tempt a naïve victim to become a hate monger. In support of this, Kteily and colleagues noted that the perception that members of an outgroup dehumanize your group (i.e., meta-dehumanization) can cause one to dehumanize the dehumanizers (2016). Ethnic-based hate speech makes some people start to believe that the language mirrors the thoughts and social attitudes of a given group to which the offender belongs. By the same token, exposure to hate speech on social media helps the expression to be self-evident and desensitizes labelers (Lakoff & Johnson, 1980; Soral et al., 2018). It is also likely that once they started exchanging badmouth with members of the opposite camp, the force of social facilitation will help poachers to engage in one-upmanship—to show how

smart they were at offending their targets. Therefore, taking proactive measures such as reducing exposure and forearming users is helpful to prevent the development of prejudice and hatred.

While trolls contribute to polarization by offending outgroups or defiant ingroups, pacemakers encourage trolls and their supporters. It is more likely that trolls hardly exist on media platforms without the supportive role of pacemakers. Pacemaking, in essence, attracts like-minded people. The attitudinal outcome discussion within like-minded individuals is more problematic. Echoing this, Sunstein, in a seminal book 'Going to Extremes: How Like-Minds Unite and Divide' admirably noted, 'when people find themselves in groups of like-minded types, they are especially likely to move to extremes. He adds, 'When likeminded discussion involves authorities who tell group members what to do, very bad things can happen' (2009, P. 2.).

Even though pace-makers fuel hatred, detection of hate speech from their words is problematic, as they just appreciate the troll, perhaps using few words. This makes holding them accountable for using HSDSPPE difficult. Pacemaking, particularly inflaming, shows individuals are more likely to remember negative things about outgroups (perhaps due to collective memory) during conflictual scenarios (Bar-Tal et al., 2014). Cognitively accessible information, such as stereotypes, appears to play a role in reminding inflamers what to comment on to fuel the debate. Furthermore, groups appear to have collective memory, as one toxic speech reminds social media users of another topic in the same domain. In support of this, Bar-Tal said in-groups contribute to the collective memories of the group as a whole (2000).

Consistent with social identity theory (Tajfel & Turner, 2004) and shared reality theory (2009; Jost et al., 2007; Hardin & Higgins, 1996), I contend that Pacemaking serves as a social support to validate the hateful expressions of the troll, which may even lead to system justification (Jost et al., 2007). When trolls know they have supporters, it is more likely that they feel safe, making it difficult to change their stance. Pacemaking such as motivating and inflaming is consistent with 'flattering' and 'that-is-not-all' known in social influence literature (Cialdini & Goldstein, 2004; Mayers, 2010).

The net effect of inflamers contributing more hate-filled topics is more likely to intensify hatred by polarizing members from both camps. For example, as trolls and pacemakers are almost on the same page (and more likely to be from the same ethnic group), members of the

target group assume the relevant outgroups are all the same in hating them (Jost et al., 2007; Hardin & Higgins, 1996). Social psychological studies show individuals' actions can be enough to elicit groups' prototypes or strengthen an already-held stereotype (Lakoff & Johnson, 1980). The online platform further publicizes these roles, influencing millions towards the whirl of hate speech (Gerbner, 1998). It has to be noted that a single troll online attracts hundreds of pacemakers through time, which could tempt members of the target group to rush into hasty generalization (which is a cognitive error) in labeling the group as 'all are the same haters' (i.e., stereotyping). In addition, the exchange of badmouth among trolls and pacemakers contributes to polarization. In this sense, I have the impression that while the idea of group polarization mainly considers like-mindedness (Yardi & Danah, 2010), the verbal fight we observed between members of two camps is equally important in understanding how polarization occurs on social media (Feilitzen, 2009). It appears to me that this is more problematic if elites or authority figures are involved, either as trolls or pacemakers.

It is interesting to have users who are concerned with peace on the platforms, even amidst disparaging conversations. Peacemakers include social media users who are willing to restrain disparaging others. Therefore, peacemakers are indicators that there are users who dare to say no to trolls within, though their number is very small. I feel that peacemakers help pacify social media platforms, provided that they outnumber trolls and pacemakers. However, based on the psychology of reactance and the evidence gathered in this study, they are not immune to receiving negative reactions and hence have to take caution to avoid negative reactions from trolls. The psychology of reactance contends that individuals will strengthen their actions or attitudes if they notice others are attempting to suppress their freedom and inner conviction (Brehm, 1966). If the stance held by the majority of ingroups is stronger, peacemakers could be considered defiant-black sheep.

It appears that Informers and Guardians could play constructive or destructive roles while reacting to hate speech posts. As such, both should take care not to offend others or lend themselves to misperceptions and misinterpretations. For instance, while helping others to be aware of some issues, they should be diplomatic so as not to offend others. Concerning the 'guarding' role, the action of documenting files could be used to accuse outgroups in the future. The documentation could be used to mobilize ingroups, call for help from outsiders, or promote the cause of ingroups. For example, some of the hate speeches we noted were

rooted in documents individuals saved at some point in the past. Besides helping ingroups in some ways, this role can also be used to fuel hatred and mobilize ingroups against outgroups in the future. Hence, while the role can help ingroups, it could be instrumental in sustaining intergroup hatred. The bystanders' role is mentioned in previous studies. For example, lurking is reported by Antin and Cheshire (2010) to imply those social media users who are known to read an online group's postings regularly but rarely participate in posting their own or sharing or reacting on others post.

Nonetheless, studies show responding to trolls is not an easy job with regard to obtaining the desired outcome. Let alone a pronounced emotional tone like anger, fear, or pride, even providing facts or exposing contradictions can backfire and attract more extremity (Lasser et al., 2023). Based on the available recommendations, counterspeech, which is the practice of responding to speech that seems harmful or offensive, could be an option to be employed judiciously (Schieb & Preuss, 2016). It can take many forms, such as challenging, debunking, or critiquing harmful speech, amplifying alternative viewpoints, providing accurate information, and fostering empathy and understanding. Miškolci et al. (2018) found that instead of responding directly to trolls, it is a useful way to reach a larger audience and provoke more counterspeech. Schieb and Preuss (2016) concluded that counterspeech can influence the original hateful speaker when counterspeakers greatly outnumber those sharing hateful messages. They also argued that a small group of counter-speakers could still be effective in cooling down moderate trolls with moderate views. Users should also be aware that their exposure to hateful sites can lead to the rise of hatred. A study by ElSherief and colleagues on 'hate speech instigators and their target', for example, found that higher visibility and participation in hate are correlated (2018). In addition, their study shows that more visible Twitter users (with more followers, retweets, and lists) are more likely to become targets of hate.

When encountering hateful posts online, it can be tempting to react and express one's opinion. However, it is important to note that responding to hate speech can make it more popular. This happens because our comment on the post easily reaches our friends and friends of friends on the platform. The spread of the hate speech post can also be facilitated because of the presence of 'spreaders of content' (Ostapeenko & Sosniuk, 2019) and 'debaters' (Cicek & Erem-Erdoğmuş, 2013). For the sake of countering hate speech, distancing oneself from trolls, such as by not following them, could help tackle hate speech

and the spread of animosity. In support of this, Gorka said that one of the methods is ignoring the haters because they want confrontation and discussing with them is futile. Among the polled population, 40 percent of those who tried this method said it was the most efficient (2019).

Even if these roles are identified from the comments users provide to hate posts, I strongly contend that the roles should be seen in a broader context of social media use during conflictual scenarios. With regard to documenting hate speech posts that target their group, for example, there are social media users (individuals or organizations) who opened a social media account specifically to document hate speeches that targeted their ethnic group. For example, ‘Tghat ትግሃት’ (@TghatMedia) and Omna Tigray (<https://omnatigray.org/>) are two of the prominent social media accounts that collect and store hate speeches made against Tigrayans during the conflict in the North. By the same token, trolls' role is also seen not only in responding to posts but also in posting hate speech content. In this sense, the roles identified in this study could likely be projected to overall social media use or engagement beyond commenting on hate speech posts.

Concerning the last research question, this study has identified how hate speech occurs. Earned reputation, trolling and poaching, and users' fallibility are the three core mechanisms through which hate speech occurs and gets intensified. With regard to earned reputation, the findings revealed that members' misdeeds (offensive episodes), liable subgroups, and hate-filled documents expose ethnic groups to hate speech. Individuals' misdeeds as a cause for hate speech imply that members make enemies of their group. The offensive actions members commit evoke anger in outgroups. As a result, decent societies (ethnic groups) are dehumanized, derogated, or mocked because of the misdeeds (including speeches) of their members. In addition, the misfortunes of members are also seen to attract hate speech to their groups. This observation implies that haters have something as a starting point for their hate speech and to justify why they label their target as such. In essence, this observation is in line with the proverb: ‘There is no smoke without fire.’ In cases of criminal or heinous actions, the hate speech targeting the whole group indicates the presence of image abusers in the group who are exposing their group to hatred.

In support of this observation, Macrae and colleagues argued that observing outgroup members behaving in stereotype-congruent behavior strengthens and justifies the already existing attitude and label, even to the extent of challenging conscious attempts to suppress

the stereotype (1995). In this sense, I assume that individuals' misdeeds can contribute both to the occurrence of hatred (a negative stereotype) and the strengthening of an already existing one. This is perhaps more pronounced because of the notion of 'group entitativity', which is the extent to which a group or collective is considered by others to be a real entity having unity, coherence, and internal organization rather than a set of independent individuals (Yzerbyt & Rogier, 2001). The phrase 'they are all the same' shows this group entitativity. The adjectives 'inherent' and 'historical' used to label groups as enemies also further support the perception that all members have the same traits. Such generalization of fault from individuals to a group is stronger when the action is done by elites or figures supposed to represent the group and when it is repeated, or is vivid.

In support of this, social psychological science on attitude formation shows that through repetition and publicity, actions by others vivid immoral acts become cognitively accessible to shape our perception of relevant others. To add fuel to this, Banaji & Bhat noted that social media contributes to the availability heuristic by popularizing vivid episodes, which leads to erroneous conclusions about a group (2022). In addition, Tajfel and Turner (1986) noted that while individuals compare their group to the relevant outgroups, it is more likely that they remember more negative information about the outgroup, and more attention is allocated to stereotype-congruent behavior.

The offensive episodes subcategory implies those actions evoke anger on the part of outgroups.

In turn, outgroups who are angered by misdeeds prefer venting out their negative emotions against the perpetrators or their associates (Berkowitz & Harmon-Jones, 2004). In support of this, Deffenbacher and colleagues show that verbal assault is among the nine different ways of expressing anger (1994). In the same vein, the general strain theory says negative emotion (e.g., anger, frustration, or stress) leads to retaliation against the perceived cause (Agnew, 2006), or hate speech against a group in this case. Therefore, even if insulting an entire ethnic group for the mistake of an individual is a cognitive error, hate mongers sometimes get evidence to justify their hate speech against their target.

While the actual immoral act of an individual (s) is the main reason to insult a group, members showing support for the wrongdoing (motivators) further the stereotype and strengthen the label through stereotypical expressions as 'they all are the same'. Indeed, some groups are ready and willing to appreciate, support, or defend their wrongdoer.

It has to be noted that when I say offensive episodes as a source for hate speech, questioning how we come to know (as a social media user) the truthfulness of the alleged wrongdoing claimed by others is important. In our digital era, technologies allow the proliferation of fake videos, audio, and photos (deepfake¹⁷) which appear real to the eyes of the majority of social media users who are presumably low in media literacy. It is, hence, logical to assume that earned reputation can take actual deeds, unproved rumors, or disinformation.

I contend that earned reputation as a means for hate speech diminishes social inhibition to disparage others and helps haters have the license to derogate their target. In addition, it also helps justify the hate speech and hence serves as a potential protective mechanism against possible cognitive dissonance that would have resulted from offending another human fellow (Festinger, 1957). In one of the mechanisms of dealing with dissonance, individuals either trivialize or leave the dissonance element so that they can live with their attitude—hate in this case. It is also likely that an earned reputation as a means of hate speech has the power to convince people to join the hate circle, perhaps because evidence is claimed to exist for the label. In this line of thought, Staub (2005, 1992) argued that perpetrators often use just-world thinking, which implies individuals and groups get what they deserve (Lerner, 1980). In a just worldview, perpetrators view their violence toward others as a response to the character, intentions, or actions of their victims.

While the path taken in generalizing from individuals' mistakes to groups is a cognitive error (called hasty generalization), the offensive episode as a means for offending ethnic groups is also congruent with the venomous group attack presented in the previous section. In venomous group attacks, trolls are seen insulting the entire group, of which an individual who is a cause for irritation is a member. It is, therefore, worth breaking this path, either by helping individuals to avoid lending their group to hate speech or by helping users not to engage in faulty generalization. By the same token, there is a need to correct the image abusers, such as by criticizing them (i.e., within-group correction). Unfortunately, although this observation calls for correcting rogues within one's group to save the group's reputation, individuals rarely condemn their wrongdoers.

Because hate filled documents are mainly written and promoted by elites, they possess the ability to convince one's group while provoking and intimidating relevant outgroup. Hateful narrations (oral or written) also attract equivalent hateful narrations from outgroups, keeping

¹⁷ Convincing image, audio, and video hoaxes, mainly produced by artificial intelligence or machine learning

all in the circle of mutual accusation (Bar-Tal et al., 2014; Bliuc et al., 2017). This mainly happens as individuals selectively deploy their text while denying and discounting others', a cognitive error called confirmation bias (APA, 2015; Shiraev & Levy, 2013). Nonetheless, such narratives and counter-narratives are known to serve to deepen social divisions, increase polarization, or justify violence, and have the potential to interrupt peace processes (IFES, 2021).

In support of this view, collective memory theories propose that groups' remembrances of their past depend upon their current social situation (Rimé & Christophe, 1997). This perhaps shows that the current sociopolitical landscape in Ethiopia dictates what to remember and bring to discussions on social media. In the same vein, Volkan (2001) argued collective experiences, including traumas (real or perceived), are transmitted across generations through storytelling and myth-making, shaping identity groups' opposition to the 'other'. This is more pronounced as discussions among those who share similar ideas are more likely to endorse the dominant idea in the group and move to extremes (Sunstein, 2009).

To make matters worse, narrations by elites appear to the audience as factual and scientific discourses. This also gives such content immunity from legal criticism. For example, the Ethiopian Hate Speech Law says a message cannot be considered hate speech if it is presented as an academic study or scientific inquiry, a news report, an analysis or political critique, artistic creativity, or religious teaching (FDRE, 2019). Therefore, textual hate speech, which is sourced from books or other texts, could challenge the Ethiopian Hate Speech Law because such expressions can be disguised in scholarly discussions and pass the hate speech screening test. Nonetheless, the current study shows hate speech is somehow associated with these immunities.

Concerning attacking liable subgroups, individuals usually show support to their subgroup and engage in defending the subgroup and image abusers within, including insulting back the haters. This observation is perhaps parallel to Allport's postulation: in-groups are "psychologically primary," in the sense that familiarity, attachment, and preference for one's in-groups come before the development of attitudes toward specific out-groups (1954, p.). The current study also shows the ripple effect of attacking subgroups and cultural markers, such as cultural leaders, ethnic elites, and traditions, on intensifying hate speech. In support of the findings in this study, Malešević noted that cultural markers can be used as a potential starting device for social chaos (2004). Although it is not inherently conflictual, ethnicity has

psychological properties and discursive resources that have the potential to decant into violence and/or social mobilization (Young, in Darby & Ginty, 2008; Malešević, 2004). This is worth attending, as ethnicity is shaping politics in Ethiopia and even causing pressing problems. An important takeaway is that, contrary to what some outgroups may perceive, subgroups are liked by ingroups. This in turn shows the need to respect the group by respecting its part or subgroup. Nonetheless, if a subgroup is legally criminalized, such as TPLF, during the conflict, individuals take insulting it as the norm because the legal aspect gives them a license to do so.

The idea of poaching is similar to hatemongering, as known in the literature. Yet, this type of user is persistent and motivated to attack a certain target. Accusation, deception, irritation, and playing credibility are the main techniques poachers employ to trap their prey. I contend that poachers who are elites have more roles in enemification by virtue of their social power. Thus they can easily influence their followers through their textual hate speech and conflict-supportive narratives (Bar-Tal et al., 2014). In this analysis, poachers appear to be effective in achieving their goal in terms of irritating their target, evoking a similar response, and allowing their content to go viral. This is partly because of the readiness and willingness of social media users to fall prey to the main trolls. By the same token, this study has noticed the toxic dialogue going on among social media users geared towards irritating one another. In addition, such irritation as a mechanism is also evident in sole of the jokes that exposes a group to devaluation. Partly, some of the viral ‘humorous’ ethnic jokes appear to be deployed both to express stereotypes and to earn money from platforms. It is, therefore, reasonable to assume there is hate speech meant for sale. Equally important to note is the emergence of ethnoreligious hate speech-simultaneously attacking an ethnic group and a religious institute, on the assumption that the two are closely linked or are the same. This observation, nonetheless, is specific to simultaneously attacking ethnic Amhara and the Ethiopian Orthodox church, demonizing and portraying them as enemy.

Accusations are more likely to escalate in the presence of real or imagined intergroup conflict when inter-group comparison is activated, present ingroups as the sole victims of intergroup relations, and call ingroups for patriotism and unity (Bar-Tal et al., 2014), a phenomenon that I believed to be present in the current Ethiopia. Accusation in the context of rival ethnic narratives, for example, easily leads to mutual accusation and accusation in a mirror (Marcus, 2012), which in turn leads to offline violence. This theory assumes victimization is a risk factor for offending others. In this sense, prey can be angered at the hate monger and become

a troll against their attacker (i.e., a vicious circle of hate speech). Needless to say, offenders' routine tasks, such as offending others, attract others not only to offend them but also their groups (Agnew, 2006). Parallel to actual conflict scenarios where perpetrators selectively attack their enemies (Bruce, 1997), this study shows that poachers on social media have ready-made 'enemies' to accuse, blame, and attribute their misfortunes. In this sense, I do support the HSDSPPE in focusing on those trolls with more than 5000 followers. Yet, I believe this figure ignores some more offenders with less social media followers. I would rather recommend the law focus on those elites and/or vocal extremists who consistently attack others (i.e., poachers), irrespective of the number of followers they have. Of course, there is a need to help users be cautious in following elites so as not to be dragged in hate the speech circle. A study has shown that users who are loyal to reliable sources use, on average, more toxic language than their counterparts (Cinelli et al., 2021). This challenges peace if the sources of information for users are hate mongers. The problem is even worse when social media is used mainly by users/followers who are gullible, low in diversity tolerance, and low in media literacy. In support of this, Kiefe succinctly noted that the internet and social media are '...at the hands of minors and hate mongers; the uncivilized internet is known for the unpalatable rogue web culture of spreading hatred and disinformation' (2017, P.1). Therefore, helping users to develop healthy skepticism while surfing the net and maintaining reasonable distance from blindly believing their source is highly needed.

In support of the findings of the current study, the available researches collectively suggest that political elites play a significant role in intergroup hatred, through using violent rhetoric that mobilize antagonism towards outgroups (Bar-Tal et al., 2014; Bradshaw & Howard, 2019; Carment, 2007; Caspersen, 2008; Glaeser, 2002). Benesch (2014), for example, found that the likelihood of violence emanating from hate speech is greater if the speech is made by high-profile individuals. In apparently interpersonal affairs, denigrating individuals, particularly elites, can be extrapolated to the intergroup level. This is perhaps because of the power of social identity. In support of this, Malesevic (2004), in his book 'The Sociology of Ethnicity', argued that one of the functions of ethnicity is to view members as extended families. This observation also helps elaborate the literature in such a way that instead of treating interpersonal versus intergroup hatred, there is a bridge that denigrating individuals in apparently interpersonal affairs can be extrapolated and implicated at the intergroup level.

Disagreement observed over ‘Ethiopian History’ also contributes to hate speech, which tempted me to perceive this context as Pandora’s Box¹⁸. In this study, the analysis shows that toxic narratives in the form of ‘history’ are used by elites and echoed by ordinary social media users to accuse, defame, and irritate outgroups. Competition over resources is also highly related to Ethiopian history. In support of this observation, social psychological science suggests that antagonistic intergroup behavior is motivated by realistic intergroup conflict (Sherif, 1966) and by gains for one’s social identity (Tajfel, 1982; Tajfel & Turner, 1979).

This study also identified that failure to appreciate or tolerate diversity, as seen in derogating cultural markers (e.g., Aba-geda and Irreecha, the Geda system, Qeero, Fano, TPLF), has contributed to the rise of hate speech in Ethiopia. The ‘ripple effect’ of attacking such ‘sensitive parts’ of groups is evident in fierce verbal fights between members of groups in a way that pollutes the platform. In support of this view, Malesevic (2004) argued that any cultural marker can be used as a potential starting device for social closure. This observation is parallel to the vast literature on multiculturalism, diversity tolerance, and cultural intelligence.

These contexts are important as they may be unique in Ethiopia or during the data collection period with unique socio-historical contexts and media environments. It has to be noted that the findings of this study could be idiosyncratic to Ethiopia and should be carefully generalized. Therefore, the types of roles played by social media users, the types of hate speech, and the mechanisms identified in this study should be understood within these contexts.

As to my knowledge, I believe that this study has some unique contributions as well as supportive evidences from previous studies. While categories and concepts (even named differently) confirm previous findings, the study has also generated new concepts and categories and hence can contribute to the existing literature in the area. The bitter takeaway is that, weaponization of public discourse based mainly on identity politics is a common issue in the ethnic-based hate speech in Ethiopia. Equally important to note is the various hate speech expressions framed by elites are also shared by ordinary social media users, making those hateful expressions familiar on social media platforms, and perhaps offline.

¹⁸ In Greek mythology, a box given to a woman buy gods; when she opened it all misfortunes came out of it.

I also believe that the findings of this dissertation are highly relevant from a theoretical point of view and practical aspects to understand and tackle hate speech in Ethiopia. For example, concepts such as earned reputation and its three subthemes and poaching help theorize the causes of hate speech and contribute to advancing the literature in the area of hate speech and intergroup hatred. They also help mitigate the spread of hate speech on social media such as by correcting the image abusers that expose groups to hate speech as well as by being attentive to the techniques poachers employ to disseminate hate speech.

With regard to the robustness of this study, I strongly believe that this work has passed through a rigorous research process and generated quality concepts which help both academicians as well as practitioners in the area of hate speech and related areas. It is my firm belief that findings of this study have conceptual density; that is, categories are theoretically dense, richly dimensional, and tightly linked with one another, qualities expected from GT (Corbin & Strauss, 1990, P. 18–19). More importantly, due to adopting GT, this study has achieved the theoretical abstractions necessary to elevate conceptual categories beyond descriptions and extend the psychological usefulness of concepts- quality measure for the robustness of GT studies (Birks & Mills, 2011).

The followings, nonetheless, are the main limitations I noticed in this study: First, because the study focuses on mainly comments in written form, the findings may have missed additional roles that might be inferred from reactions such as emojis, likes, dislikes, shares, etc. that users react to posts. Therefore, specific reactions (other than comments in text form) could have helped generate additional roles or substantiate or supplement the roles identified in the current study.

Second, while a few high-profile name accounts included in this study are checked and proved to belong to the name they bear, there could exist fake accounts whose contents are included in the data of this study. In similar way, for some of the hate speech contents analyzed in this study, especially the hate comments given to main posts, it was not possible to specifically figure out who the actual sources of the hate content are and what their true motive is. This paper also understands that some users may pretend as if they belong to an ethnic group to which they don't in reality, and spread hate speech by concealing their true ethnic affiliation. I assumed their ethnic affiliations at face value, such as by looking at their name and the ideas they promoted. In related manner, this paper didn't rule out whether artificial intelligence, AI (deepfake) such as bots contributed to the contents analyzed,

especially the videos. Some of the contents of social media could be the result of social media bots, which are automated programs used to engage in social media, such as to respond in some ways. Therefore, understanding the finding of this study should take these issues into account. Nonetheless, whoever posts hate speech content online or provides negative comments has the power to stir social media with hate, which in turn can cause polarization as well as negative offline consequence. What is studied in this dissertation is publicly available hate speech contents and comments given to them, irrespective of who the actual source is. In this sense, I strongly contend that no ethnic group should be held accountable to the hate speech content analyzed in this work. Rather the contents analyzed here are individuals' production, with the possibility of polluting the social media with hate and hence contributing to societal polarization.

Third, based on recommendations on data analysis in grounded theory, it is necessary to have some views from participants (social media data sources) regarding the codes that emerged and concepts or categories that developed (Branthwaite & Patterson, 2011). Yet, the researcher couldn't interview actual hate speech producers because doing so would be a challenging task from practical point of view. The researcher believes that failure to take in the ideas of trolls and poachers remain a limitation in this study. To fill this gap, nonetheless, participants who have fairly good experience with hate speech on social media in Ethiopia have been contacted and interviewed.

Fourth, this study relied on two languages: English and Amharic. Even if Tigrigna and Afan Oromo should have been included, as much hate speech was also disseminated in these language, the researcher couldn't understand Tigregna language. Hence, in order to balance, Amharic and English languages were considered as Amharic is almost the most widely spoken in Ethiopia while English is a language that educated people prefer to disseminate their ideas to their audience. Nonetheless, members could communicate ideas including hate speech using their language within closed system, it would have been better to include these two languages in this study. Therefore, missing ideas written in these two languages remain the limitation of this dissertation work.

5.2 Conclusions

This study aimed to explore and explain the manifestations (types) and mechanisms involved in the occurrence and propagation of ethnic-based hate speech. The study primarily used social media archives. While some hate speech content is created by account holders (hot spots), a significant portion of the hate speech content analyzed in this study came from a diverse group of social media users, including high-profile figures, religious teachers, scholars, activists, and ordinary citizens, including Ethiopians in the diaspora. The collected social media contents were analyzed rigorously using GT approach.

This study identified four types of ethnic-based hate speeches: dehumanization, enemification, devaluation, and attacking and repelling. In relation to these findings, the current study also identified worse-than-demon and less-than-animal expressions, which are both extreme in their tone as well as unique to this study. Second, this study has identified five main ways in which users react to hate speech posts: trolling, pacemaking, peacemaking, guarding, and informing. Concerning the last research question, three main categories are identified that help explain how hate speech is occurring: earned reputation, poaching, and users' fallibility. This paper contends that earned reputation, selective sensitivity, and poaching are new concepts introduced into the hate speech literature. Based on the context and the tone of the expressions of the hate speeches analyzed in this study, I contend that they are of a higher magnitude capable of causing massive atrocities. The less-than-animals and 'worse-than-demon' expressions made by high-profile names and broadcasted publicly makes the hate speech scenario very dangerous.

The study showed that identity politics is at the heart of hate speech on social media. More specifically, the majority of the hate speeches analyzed are framed in the context of history and narrations, competition over resources like land, the conflict in the North, failure to uphold law and order, and elites' destructive role.

Mistrust in politics, stereotypes, actual offensive deeds, disinformation and rumors, conspiracy, elite-supported conflict supportive narratives, and hateful texts (some disguised in historical facts), as well as irresponsible use of social media by gullible users and hatemongers, have contributed to the rise of hate speech in Ethiopia. It is also important to note that Ethiopians in the diaspora contribute to the hate speeches analyzed in this study.

In this study, three ethnic groups are mainly implicated in the findings: ethnic Oromo, Amhara, and Tigray. I strongly believe that hate speech is a burden to these three ethnic groups as well as the remaining ethnic groups in Ethiopia. More importantly, hate speech, as

seen in this study, has the potential to contribute to perverse social capital and inter group interactions that result in hampering social relations and reducing possible social benefits of interethnic relations. The excerpts presented as evidence are selected merely to understand the situation through illustrating the concepts and categories under discussion and, hence, shouldn't be used to fuel further conflict and hate speech. Furthermore, I reiterate that the examples presented in this monograph are productions of individuals who behave on social media irresponsibly.

Chapter Six

Implication and Recommendations

As hate speech is the mechanism by which people produce their enemies, which ultimately could lead to actual violence sooner or later, informed interventions should be put in place. For this, I believe that addressing hate speech should start with acknowledging that while freedom of expression is a basic human right, the rise of social media users has significantly expanded the channels for creating and spreading hate speech in today's Ethiopia. This situation has already exposed Ethiopians to threatening and degrading situation. I also believe that the desired outcome of measures to combat hate speech in Ethiopia must identify and address the underlying structural, historical, situational, and individual-level factors that fuel the issue. Moreover, countering hate speech should involve all relevant stakeholders and actors in one or another way. Based on the findings, I recommend the following to nip hate speech in the bud before we go far in the continuum of destruction:

I. For the Government, Political Leaders, and Ethnic Elites

The government should play the following roles to counteract hate speech: First, this study has shown that the various forms of identity-based violence perpetrated in different corners of the country, such as the eviction of people from their residential areas (IDPs), extrajudicial killings, barbaric and vandalism acts, and the unappealing public appearance and speech of political elites, have exposed ethnic groups to hate speech and negative stereotypes. For example, the finding showed that when certain criminal or heinous acts are committed in a particular place by an individual or a few individuals, the ethnic group with claimed ownership of the place where the act is committed is bitterly insulted, including with 'worse-than-demon' and 'less-than-animal' expressions. I contend that this kind of hate speech occurrence is partly attributed to the failure of the government to uphold law and order, which in turn allowed vandals to offend others in a way that exposed their group to hate speech. Therefore, I strongly recommend that the government, through its legal means, enforce law and order so that hate speech against an ethnic group can be significantly reduced. In addition, anyone who represents the government should act professionally and responsibly to prevent hate speech against the government as well as the ethnic group one claim to belong to. The data also showed politicians' unappealing public speech could also be documented as evidence to fuel intergroup hatred. Hence, it is reasonable to recommend that the government correct its public figures to responsibly behave publicly. The government

should also hold leaders accountable for the hate speech they produce and disseminate. This action, nonetheless, requires stronger political commitment and will.

This study has also shown that decent societies are devalued because of the wrong or offending actions of ethnic elites and ethnic representatives. Therefore, any individual or subgroup that claims to or is supposed to represent an ethnic group should refrain from publicly committing offenses both online and offline and promptly speak out against violence and hate speech. I also recommend the need to publicly apologize for the hate speech disseminated before. This works for both political leaders and other ethnic elites.

Second, failure to remove hate speech from social media platforms contributes to offline conflict, the rise of trolls (increased exposure), and helps document hate speech contents for future destructive roles (textual hate speech). In this sense, many of the hate speech contents used in this dissertation are still available on social media, nearly three or more years after they were posted. This shows social media companies do not remove hate speech content per their community standards. This calls for the need to collaborate with and enforce them so that they act responsibly, per their community standards. Based on Stanton's (1998) and Benesch's (2014) recommendation, I suggest media platforms taking preventative measures to vigorously protest the use of dehumanizing words as well as calls for attacking and repelling types of expressions.

With regard to HSDSPPE, a significant portion of hate speech content, in this study, are found to be disguised within humor or jock, historical discussions, as well as religious teaching (ethnoreligious hate speech). Therefore, I recommend reconsidering the justification used to rule out the legality of hate speech framed within these contexts: artistic, scientific, and religious teaching. In addition, even if focusing on vocal extremists can help counter hate speech, the 5000-follower limit lacks justification, as haters with less than 5000 followers, as well as non-social media users can still share hate speeches with equal or more negative consequences. I, therefore, recommend the law to focus on those elites and/or vocal extremists who consistently attack others (i.e., poachers), even if they are followed by a smaller number of social media users or are inactive on the platforms. I also recommend the HSDSPPE to consider the four types of hate speeches identified in this study and, hence, redefine its hate speech conceptualization.

II. Partners Working in the area of Peacebuilding

Media companies (i.e., television and radio platforms) and organizations working on conflict resolution or peacebuilding, including Ethiopian National Dialogue Commission, should also discharge the following roles: Understanding and dealing with interethnic conflict in Ethiopia should consider online hate speech scenario. More specifically, partners working on peacebuilding, such as the Ethiopians National Dialogue commission, should give due attention to hate speeches shared on various social media platforms. It is worth noting that archived online hate speech content and hateful texts are more likely to affect the present and future generations through contributing to polarization. Therefore, the explosive remnants on social media (archived hate speech contents) and textual content should be ‘demined, such as through removing them or dealing with them for mutual benefit. This will help avoid the use of such toxic content to fuel conflict in the future.

Because narratives have the power to generate violence, policymakers and stakeholders working for peace must constantly look for windows of opportunity to interact with national and subnational narratives to redirect a society's path away from narratives that promote conflict and the negative outcomes they may produce. Equally important is the need to work aggressively to enhance diversity tolerance. There is also a need to design a strategic communication intervention to counter hate speech. For example, an awareness-raising program that aims to educate people about its nature, means, consequences, and countermeasures should be put in place using mass media and social media platforms.

Partners should also engage in helping social media users develop the skills needed to peacefully navigate social media platforms. One way to neutralize the destructive role of poachers is to understand their presence and the possible strategies they employ to trap other social media users. As the saying ‘forewarned is forearmed’ entails, helping social media users enhance their media literacy and diversity tolerance and urging them to use platforms responsibly could help reduce hate speech. Interventions should also focus on helping users to appraise and constructively handle competing historical narratives (e.g., to appreciate how narratives hurt and help their society). Equally important is including cyber safety education in the school curriculum, peace education, critical thinking skills, social and emotional intelligence skills, as well as enhancing responsibility and civility in using social media to combat stereotypes. All relevant stakeholders should deal with contested but ‘licensed’ terms, such as ‘Oromuma’ and ‘Nefitegna’, which are nearly freely used to offend groups. I,

therefore, recommend reaching for mutual respect and understanding regarding such emotive terms.

III. For Social Media Users

Users should responsibly use the media. They should be aware that reacting to hate speech could promote the content, either directly or indirectly. Abstaining from commenting, providing constructive comments or reporting the post to legal authorities and media companies should be done judiciously. They also better examine the context in which the hate speech is made, who the trolls and their supporters are, and the readiness and capacity of the parties involved in the comment line before deciding whether to respond to a hate speech post. They should also carefully examine their media literacy skills, diversity tolerance, and emotional intelligence skills, which are influential in deciding the outcome of interacting with hateful content. Self-awareness in these skill domains could help them take proper action while surfing the net. This study has shown individuals usually support their members or abstain from criticizing their members, even if they are offenders. Nonetheless, for the sake of the reputation of their group, members of a group should criticize those individuals who engage in attacking others as well as abstain from supporting the trolls within. In addition, they should take care when derogating subgroups and ethnic markers, lest they could offend the entire group.

IV. For Researchers

In terms of future research, it would be useful to extend the current findings by examining the following: First, this dissertation merely explored the types of hate speech and the roles of users. Counting each category was not the intention of this work. Nonetheless, as counting them would help us prioritize intervention areas, I recommend quantifying the types of hate speeches and the roles played by social media users. Second, this paper explored types of hate speech and the mechanisms involved in their occurrence. Nonetheless, there is no evidence regarding how those hate speeches contributed to actual violence. Therefore, it is better to examine to what extent online hate speech contributes to the offline conflict in Ethiopia, especially during conflictual scenario. This helps us identify the level of the link.

Third, there is evidence showing that ethnic-based hate speeches are infused with religion. For example, hate speech against Amharas is sometimes mixed with that of the Ethiopian Orthodox Church. Therefore, I recommend independently investigating ethnoreligious hate speech in Ethiopia. Fourth, Ethiopian researchers on conflict and peace should allocate

reasonable attention and resources to include social media. Therefore, studies on conflict and peacebuilding in Ethiopia should provide a fair representation of social media sites in their studies. Fifth, because the role ‘peacemaking’ by social media users is helpful to cool verbal fights online, I recommend further studies to explore how they are reacted to, ways to maximize their impact, and identify the personality traits, motivations, and worldviews that likely underpin this role for future learning. Last but not least, this study has identified severe forms of ethnic-based hate speech in highly sensitive contexts (i.e., the conflict in the north). While recognizing the possible role of hate speech in contributing to various identity-based attacks and vandalism actions observed during the data collection period, I still feel that some buffering factors were there that prevented the direct consequences of the calls made to attack members of targeted groups. Hence, I recommend future studies to investigate possible protective factors that can minimize the link between online hate speech and offline conflict in Ethiopia, especially during such conflictual scenarios. This would help us in capitalizing on those factors, while at the same time countering the actual hateful expressions.

References

- Abay M. (December 23, 2021). *ሸር ተጋሩ ካብ ሰብ ዘይተፈጠሩ እዮም* (Video file). Retrieved from <https://www.facebook.com/watch/?v=1173057789888706>
- Abraha, M. (2019). *Mapping Online Hate Speech among Ethiopians: In Case of Facebook, Twitter and YouTube*. [Unpublished Masters Thesis]. Addis Ababa University. Available at: <http://etd.aau.edu.et/handle/123456789/18880>
- Adams, W. C. (2015). *Conducting Semi-Structured Interviews. Handbook of Practical Program Evaluation*, 492–505. doi:10.1002/9781119171386.ch19
- Adler, E.R., Hebel-Sela, S., Leshem, O.A., Levy, J., & Halperin, E. (2021). *A social Virus: Intergroup Dehumanization and Unwillingness to Aid Amidst COVID-19 – Who are the main targets?* International Journal of Intercultural Relations, 86, 109 - 121.
- Agnew, R. (2006). *Pressured into Crime: An overview of General Strain Theory*. Los Angeles: Roxbury.
- Aldiabat, K. M & Navenec, C. L. (2018). *Data Saturation: The Mysterious Step in Grounded Theory Method*. The qualitative Report, (23) 1.
- Allport, G. W. (1954). *The Nature of Prejudice*. MA. Addison-Wesley.
- Amdur, R. (1980). Scanlon on Freedom of Expression. *Philosophy and Public Affairs*, 9(3):287-300.
- American Psychological Association (2015). *Dictionary of Psychology* (2nd ed.). Author. Washington Dc.
- American Psychological Association (2020). *Publication Manual of the American Psychological Association: (7th ed.)*. Author. Washington Dc.
- Amhara Communication (September 1, 2021). *በሐረ ድሬርቃ ግንባር የሚገኙ የፋኖ አባላት ጀብድ* (Video file). Accessed from <https://www.facebook.com/watch/?v=212084204286251>
- Anstey, M., Zartman, W., & Meerts, P., (2012). *The Slippery Slope to Genocide: Reducing Identity Conflicts and Preventing Mass Murder*. New York. Oxford.
- Anti-Defamation League (2005). *Pyramid of Hate*. Accessed on 20 Jan 2023 from https://media.al.com/business_impact/other/Pyramid-of-Hate.pdf
- Article-19 (2015). *Hate Speech Explained. A Toolkit*. London. Retrived on june 6, 2019 from <https://www.article19.org/data/files/medialibrary/38231/'Hate-Speech'-Explained---A-Toolkit-%282015-Edition%29.pdf>

- Ayalew, A.G. (2022). *The Disparity of Ethiopia's Hate Speech and Disinformation Prevention and Suppression Proclamation in Light of International Human Rights Standards*. *Hawassa University Journal of Law*, 6, 75–100.
- Ayalew, Y.E (2020). *Defining 'Hate Speech' under the Hate Speech Suppression Proclamation in Ethiopia: A Sisyphean eExercise?* *Ethiopian Human Rights Law Series V*. (12)
- Babchuk, W. A. (2011). *Grounded Theory as a "Family of Methods: A Genealogical Analysis to Guide Research*. *US-China Education Review*, A3, 383-388.
- Baker, C.E. (1997). *Harm, Liberty, and Free Speech*. *Southern Calif. Law Rev.* 70:979–1020
- Barberá, P. (2020). *Social Media, Echo Chambers, and Political Polarization*. In N. Persily & J. Tucker (Eds.), *Social Media and Democracy: The State of the Field, Prospects for Reform* (SSRC Anxieties of Democracy, pp. 34-55). Cambridge University Press.
- Bar-Tal, D. (1993). *Patriotism as Fundamental Beliefs of Group Members*. *Politics & the Individual*, 3(2), 45–62.
- Bar-Tal, D., Graumann, C.F., Kruglanski, A.W., & Stroebe W. (EDs.) (1989). *Stereotyping and Prejudice: Changing Conceptions*. New York. Springer Science.
- Bar-Tal, D., Oren, N., & Nets-Zehngut, R. (2014). *Sociopsychological Analysis of Conflict-Supporting Narratives: A general framework*. *Journal of Peace Research*, 51(5), 662–675.
- Baumeister, R. F. & Finkel, E. J. (2010) (Eds). *Advanced Social Psychology: The State of the Science*. New York. Oxford University Press
- Baumeister, R. F., & Leary, M. R. (1995). *The Need to Belong: Desire for Interpersonal Attachments as a Fundamental Human Motivation*. *Psychological Bulletin*, 117(3), 497–529. <https://doi.org/10.1037/0033-2909.117.3.497>
- Baumeister, R. F., Smart, L., & Boden, J. M. (1996). *Relation of Threatened Egotism to Violence and Aggression: The Dark Side of High Self-Esteem*. *Psychological Review*, 103(1), 5–33. <https://doi.org/10.1037/0033-295X.103.1.5>
- Baumeister, R.F., & Vohs, k. D., (EDs.) (2007). *Encyclopedia of Social Psychology* (Vol.1). Los Angeles. Sage.

- Beck, A. T (1999). *Prisoners of Hate: The Cognitive Basis of Anger, Hostility, and Violence*. HarperCollins. Springer
- Bem, D. J. (1972). *Self-Perception Theory*. *Advances in Experimental Social Psychology*, 1–62. doi:10.1016/s0065-2601(08)60024-6
- Benesch, S. (2014). *Countering Dangerous Speech: New Ideas for Genocide Prevention*. Accessed from <https://www.ushmm.org/m/pdfs/20140212-benesch-countering-dangerous-speech.pdf>
- Benesch, S., Glavinic, T., Manion, S., & Buerger, C.. (2018). *Dangerous Speech: A Practical Guide*. Accessed online on June 12, 2020 from https://www.researchgate.net/publication/330684579_Dangerous_Speech_A_Practical_Guide/link/5d80f37b299bf10c1ab3e1c4/download
- Ben-Ze'ev, A. (2000). *The Subtlety of Emotions*. The MIT Press. London
- Ben-Ze'ev, A. (1992). *Anger and Hate*. *Journal of Social Philosophy*, 23(2), 85–110. doi:10.1111/j.1467-9833.1992.tb00295.x
- Bergh, R., Akrami, N., Sidanius, J., & Sibley, G. (2016). *Is Group Membership Necessary for Understanding Generalized Prejudice? A re-evaluation of why prejudices are interrelated*. *JPSP*, 111(3), 367–395. <https://doi.org/10.1037/pspi0000064>
- Berkowitz, L.J., & Harmon-Jones, E. (2004). *Toward an Understanding of the Determinants of Anger*. *Emotion*, 4 2, 107-30
- Birks , M. & Mills, E.J. (2011). *Grounded Theory: A Practical Guide*. Los Angeles, CA: Sage
- Bizumic, B. & Duckitt, J. (2007). *Varieties of Group Self-Centeredness and Dislike of Specific Other*. *Basic and Applied Social Psychology*, 29:2, 195-202
- Bossetta, M. (2018). *The Weaponization of Social Media: Spear Phishing and Cyberattacks on Democracy*. *Journal of International Affairs*, 71(1.5), 97-106.
- Bradshaw, S. & Howard. P. N. (2019). *The Global Disinformation Disorder: Global Inventory of Organized Social Media Manipulation*. Working Paper. Oxford, UK: retrieved from <https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/09/CyberTroop-Report19.pdf>
- Brady, W., McLoughlin K., Doan, N., & Crockett, M. (2021). *How Social Learning Amplifies Moral Outrage Expression in Online Social Networks*. *Science Advances*. V(7).
- Braun, V. & Clarke, V. (2006). *Using Thematic Analysis in Psychology: Qualitative Research in Psychology*, 3 (2). pp. 77-101.

- Brehm, J. W. (1966). *A Theory of Psychological Reactance*. New York: Academic Press.
- Brennan, W. (1995). *Dehumanizing the Vulnerable: When Word Games Take Lives*. Chicago: Loyola University Press.
- Brewer, M. B. (1999). *The Psychology of Prejudice: In-group Love or Out-group Hate?* *Journal of Social Issues*, 55, 429–444.
- Brewer, M. B., & Gardner, W. (1996). *Who is this "We"? Levels of Collective Identity and Self Representations*. *Journal of Personality and Social Psychology*, 71(1), 83–93. <https://doi.org/10.1037/0022-3514.71.1.83>
- Brindle, A. (2016). *The Language of Hate: A corpus Linguistic Analysis of White Supremacist Language*. New York: Routledge.
- Brison, S. J. (1998). The Autonomy Defense of Free Speech. *Ethics*, 108(2), 312–339. doi.org/10.1086/233807
- British Psychological Society (2013). *Ethics Guidelines for Internet-mediated Research*. Available at: <https://www1.bps.org.uk/system/files/Public%20files/inf206-guidelines-forinternet-mediated-research.pdf> (accessed 25 May 2018).
- Brogaard, B. (2020). *Hatred: Understanding Our Most Dangerous Emotion*. Oxford.
- Brown, A. (2017). *What is Hate Speech?* Part 1: The Myth of Hate. *Law Philos.* 36(4):419–68
- Bruce, S. (1997). *Victim Selection in Ethnic Conflict: Motives and Attitudes in Irish Republicanism*. *Terrorism and Political Violence*, 9(1), 56–71.
- Brudholm, T., & Johnsen, B.S.(2018). *Hate, Politics, Law: Critical Perspective on Combating Hate*. Oxford University Press
- Bryant, A., & Charmaz, K. (2012). *Grounded Theory and Psychological Research*. <https://doi.org/10.1037/13620-003>
- Calogero, R. (2012). *Objectification Theory, Self-objectification, and Body Image*. In: Cash, Thomas, ed. *Encyclopedia of Body Image and Human Appearance*. Academic Press, pp. 574-580.
- Carlson, C. & Rousselle, H. (2020). *Report and Repeat: Investigating Facebook's Hate Speech Removal Process*. *First Monday*, 25(2). <https://doi.org/10.5210/fm.v25i2.10288>
- Carment, D. (2007). *Exploiting Ethnicity: Political Elites and Domestic Conflict*. *Harvard International Review*, 28(4).

- Caspersen, N. (2008). *Intragroup Divisions in Ethnic Conflicts: From Popular Grievances to Power Struggles. Nationalism and Ethnic Politics*, 14(2), 239–265.
<https://doi.org/10.1080/13537110801985070>
- Cervone, C., Augoustinos, M., & Maass, A. (2020). *The Language of Derogation and Hate: Functions, Consequences, and Reappropriation. Journal of Language and Social Psychology*, 40, 80 - 101.
- Chaffey, D. (2022). *Global Social Media Statistics Research Summary 2022. smart insights*. Available at: <https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research>
- Chakraborty, T., Kumar A., Upadhyay P., Dwivedi Y.K. (2021). *Link between Social Distancing, Cognitive Dissonance, and Social Networking Site Usage Intensity: A country-level study during the COVID-19 outbreak. Internet Res*; 31:419–456.
- Charmaz, K. (2017). *The Power of Constructivist Grounded Theory for Critical Inquiry. Qualitative Inquiry* 23 (1): 34–45. doi:10.1177/1077800416657105
- Charmaz, K. (2006). *Constructing Grounded Theory: A practical Guide Through Qualitative Analysis*. Sage. London
- Charmaz, K. (2000). *Grounded Theory: Objectivist and Constructivist Methods*. 2nd Edition, Sage Publications, London.
- Charmaz, K. & Thornberg, R. (2020): *The Pursuit of Quality in Grounded Theory, Qualitative Research in Psychology*, DOI: 10.1080/14780887.2020.1780357
- Charmaz, K. & Mitchell, R. (2001). *Grounded Theory in Ethnography*. In *Handbook of Ethnography* (pp. 160-174). SAGE Publications.
<https://doi.org/10.4135/9781848608337>
- Chavan, S., & Shylaja, S. (2015). *Machine Learning Approach for Detection of Cyber-Aggressive Comments by Peers on Social Media Network. International Conference on Advances in Computing, Communications and Informatics*, 2354-2358.
- Cheng, J., Bernstein, M., Danescu-Niculescu-Mizil, C., and Leskovec, J. (2017). *Anyone Can Become a Troll*: in Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. New York, doi: 10.1145/2998181.2998213

- Chin, J. L. (2004). *The Psychology of Prejudice and Discrimination: Vol. 1. Racism in America*. London: Praeger
- Cho, J.Y., & Lee, E.(2014). *Reducing Confusion about Grounded Theory and Qualitative Content Analysis: Similarities and Differences*. *The Qualitative Report*. V.19 (32). pp 1-20. <https://doi.org/10.46743/2160-3715/2014.1028>
- Cialdini, R. & Goldstein, N. (2004). *Social Influence: Compliance and Conformity*. *Annual review of psychology*. 55. 591-621. 10.1146/annurev.psych.55.090902.142015.
- Cialdini, R. (2009). *Influence: The Psychology of Persuasion*. HarperCollins.
- Çiçek, M., & Erem-Erdoğan, I.(2013). *Social Media Marketing: Exploring the User Typology in Turkey*. *International Journal of Technology Marketing*, 8(3), 254.
- Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). *The Echo Chamber Effect on Social Media*. *Proceedings of the National Academy of Sciences of the United States of America*, 118(9), 1–8. doi.org/10.1073/pnas.2023301118
- Citron, D.K., & Norton, H.L. (2011). *Intermediaries and Hate Speech: Fostering Digital Citizenship for Our Information Age*. *Boston University Law Review*, 91, 1435. Available at: https://scholarship.law.bu.edu/faculty_scholarship/614
- Coles, B.A., & West, M.L. (2016). *Trolling the Trolls: Online Forum Users Constructions of the Nature and Properties of Trolling*. *Comput. Hum. Behav.*, 60, 233-244.
- Corbin, J., & Strauss, A. (1990). *Grounded Theory Research: Procedures, Canons, and Evaluative Criteria*. *Qualitative Sociology* 13 (1):3–21. doi:10.1007/BF00988593.
- Corlett, A. J. & Francescotti, R. (2002). *Foundations of a Theory of Hate Speech*. *The Wayne Law Review* 48 (2002): 1071-1100, p. 1083.
- Council of Europe (2023). *Study on Preventing and Combating Hate Speech in Times of Crisis*. Accessed from <https://rm.coe.int/-study-on-preventing-and-combating-hate-speech-in-times-of-crisis/1680ad393b>
- Council of Europe (2021). *Racism, Intolerance, Hate Speech: A compilation of texts adopted by the Parliamentary Assembly of the Council of Europe*. Available at <https://assembly.coe.int/LifeRay/EGA/NoHate/RacismIntoleranceHateSpeech-EN.pdf>

- Council of Europe (2020). *Combating Hate Speech*. Author. Accessed from <https://rm.coe.int/prems-083822-gbr-2018-recommendation-on-combating-hate-speech-memorand/1680a710c9>
- Council of Europe (2017). *Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making*.
- Council of Europe (2009). *Manual on Hate Speech*. Author.
- Covell, K. (1996). *National and Gender Differences in Adolescents' War Attitudes*. *International Journal of Behavioral Development*, 19(4), 871–883. doi.org/10.1080/016502596385622
- Crandall, C. S., Eshleman, A., & O'Brien, L. (2002). *Social Norms and the Expression and Suppression of Prejudice: The struggle for internalization*. *Journal of Personality and Social Psychology*, 82, 359–378. doi.org/10.1037/0022-3514.82.3.359
- Creswell, J. W. & Poth. C. N. (2018). *Qualitative Inquiry & Research Design: Choosing Among Five Approaches* (4th ed.). Sage. London
- Crises Group (n.d). What Facebook Does (and Doesn't) Have to Do with Ethiopia's Ethnic Violence. Accessed from <https://www.crisisgroup.org/africa/horn-africa/ethiopia/what-facebook-does-and-doesnt-have-do-ethiopias-ethnic-violence>.
- DataReportal (2023). DIGITAL 2023: ETHIOPIA. <https://datareportal.com/reports/digital-2023ethiopia#:~:text=There%20were%2020.86%20million%20internet,percent%20of%20the%20total%20population>
- Deffenbacher, J. L., Oetting, E. R., & Lynch, R. S. (1994). *Development of a Driving Anger Scale*. *Psychological Reports*, 74(1), 83-91. <https://doi.org/10.2466/pr0.1994.74.1.83>
- Dellagiacoma, L, Geschke, D., & Rothmund, T. (2024). *Ideological Attitudes Predicting Online Hate Speech: The Differential effects of Right-wing Authoritarianism and Social Dominance Orientation*. *Front. Soc. Psychol.* 2:1389437. doi: 10.3389/frsps.2024.13894
- Denzin, N. K. (1971). *The Logic of Naturalistic Inquiry*. *Social Forces*, 50(2), 166–182. <https://doi.org/10.2307/2576935>
- Dharmapala, D., Garoupa, N. & McAdams, R.H. (2008). *Belief in a Just World, Blaming the Victim, and Hate Crime Statutes*. University of Chicago Public Law & Legal Theory Working Paper No. 242.
- Dolinski, D.(2016). *Techniques of Social Influence: the psychology of gaining compliance*. London. Routedge.

- Dovidio, J. F., & Gaertner, S. L. (1993). *Stereotypes and Evaluative Intergroup Bias*. In D. M. Mackie & D. L. Hamilton (Eds.), *Affect, cognition, and stereotyping: Interactive processes in group perception* (p. 167–193). Academic Press.
- Dovidio, J.F., Gaertner, S.L., Pearson, A.R. & Riek, B.M. (2005). *Social Identities and Social Context: Social Attitudes and Personal Well-Being*. 10.1016/S0882-6145(05)22009-X
- Dreu, D., Gross, J., Méder, Z., Giffin, R., Prochazkova, E., Krikeb, J., & Columbus, S. (2016). *In-group Defense, Out-group Aggression, and Coordination Failures in Intergroup conflict*. *Proceedings of the National Academy of Sciences*, 113, 10524 - 10529.
- Duarte, N., Llanso, E., & Loup, A. (2018). Mixed Messages? *The Limits of Automated Social Media Content Analysis*. In Proceedings of the 1st Conference on Fairness, Accountability and Transparency, Vol. 81 (pp. 106). New York:
- Easterly, W., & Levine, R. (1997). *Africa's Growth Tragedy: Policies and Ethnic Divisions*. *The Quarterly Journal of Economics*, 112(4), 1203–1250.
- Ekman, P. (1992). *Chapter three: Basic Emotions*. in Dalglish, T. and power, M. (1999). *Handbook of cognition and emotion*. Johns Wiley and Sons.
- Elias, B. (March 7, 2022). ለአግሩ የሚጫመው ምንም አይነት ጫማ ሳይኖረው (Video file). Accessed from <https://www.facebook.com/watch/?ref=saved&v=920715091946802>
- Elias, B. (November 21, 2020). የአርበኛ ሞግ ቀራኒዮ መልዕክት ለአማራ ሰዎች "የምወድህ ህዝቤ አማራ ሆይ መልዕክት አለኝ አድምጠኝ " ይልሃል ሞግ ቀራኒዮ ዘ እምነ ነገደ አማራ (Video file). Accessed from <https://www.facebook.com/eth.zion.elias/videos/3405468382884611/>
- Ellison, N. B., Vitak, J., Gray, R., & Lampe, C. (2014). *Cultivating Social Resources on Social Network Sites: Facebook Relationship Maintenance Behaviors and their Role in Social Capital Processes*. *Journal of Computer-Mediated Communication*, 19(4), 855-870.
- Elo, S., & Kyngäs, H. (2008). *The Qualitative Content Analysis Process*. *Journal of Advanced Nursing*, 62, 107-115.
- Elo, S., Kääriäinen, M., Kanste, O., Pölkki, T., Utriainen, K., & Kyngäs, H. (2014). *Qualitative Content Analysis: A Focus on Trustworthiness*. DOI: 10.1177/2158244014522633

- ElSherief, M., Nilizadeh, S., Nguyen, D., Vigna, G., & Belding, E. (2018). *Peer to Peer Hate: Hate Speech Instigators and their Targets*. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 12, No. 1).
- European Institute of Peace (2021). *Fake News Misinformation and Hate Speech in Ethiopia: A Vulnerability Assessment*. Retrieved from <https://www.eip.org/new-publication-fake-news-misinformation-and-hate-speech-in-ethiopia-a-vulnerability-assessment/>
- Ezeibe, C. (2021). *Hate Speech and Election Violence in Nigeria*. *Journal of Asian and African Studies*, 56(4), 919-935.
- Ezeibe, C. C. & Ikeanyibe, O.M. (2017). *Ethnic Politics, Hate Speech, and Access to Political Power in Nigeria*. *Africa Today*, 63(4), 65. doi:10.2979/africatoday.63.4.04
- FDRE (2020). Federal Negarit Gazette No. 26 , March 23th 2020. *Hate-Speech-And-Disinformation-Prevention-And-Suppression-Proclamation*. Accessed from <http://www.lawethiopia.com/images/>
- Feilitzen, C. (2009). *Influences of Mediated Violence: a brief research summary*. International Clearinghouse on Children, Youth and Media. University of Gothenburg
- Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford University Press.
- Fischer, A., Halperin, E., Canetti, D., & Jasini, A. (2018). *Why We Hate*. *Emotion Review*, (10),4, P. 309–320. doi.org/10.1177/1754073917751229
- Fischer, E., & Reuber, R. (2011). *Social Interaction via New Social Media: (How) can interactions on Twitter affect effectual thinking and behavior?* *Journal of Business Venturing*, 26, 1-18.
- Fiske, S. T., & Dyer, L. M. (1985). *Structure and Development of Social Schemata: Evidence from Positive and Negative Transfer Effects*. *Journal of Personality and Social Psychology*, 48(4), 839.
- Fiske, S. T. & Taylor, S. (1984). *Social Cognition*. Cambridge, MA: Addison-Wesley.
- Fiske, S. T., & Pavelchak, M. A. (1986). *Category-based versus Piecemeal-based Affective Responses*. In Sorrentino R. M. & E. T. Higgins (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (pp. 167–203). Guilford Press.
- Fitzgibbons, R. (1986). *Cognitive and Emotional Use of Forgiveness in the Treatment of Anger*. *Psychotherapy*, 23, 629-633.
- Flore, M., Balahur, A., Podavini, A. & Verile, M. (2019). *Understanding Citizens' Vulnerabilities to Disinformation and Data-Driven Propaganda*. Publications Office of the European Union, Luxembourg, doi:10.2760/919835

- Ford, T.E., & Ferguson, M.A. (2004). *Social Consequences of Disparagement Humor: A Prejudiced Norm Theory*. *Personality and Social Psychology Review*, 8, 79 - 94.
- Forsyth, D. R. (2010). *Group Dynamic* (5th ed.). Wadsworth. Belmont.
- Fossheim, H. & Ingierd, H (2015). *Internet Research Ethics*. Available at:
<https://press.nordicopenaccess.no/index.php/noasp/catalog/view/3/1/9---1>
- Founta, A., Djouvas, C., Chatzakou, D., Leontiadis, I. Blackburn, J., Stringhini, G., Vakali, A., Sirivianos, M. , & Kourtelli, N.(2018). *Large Scale Crowdsourcing and Characterization of Twitter Abusive Behavior*.
- Frey, D. L., & Gaertner, S. L. (1986). *Helping and the Avoidance of Inappropriate Interracial Behavior: A strategy that perpetuates a nonprejudiced self-image*. *Journal of Personality and Social Psychology*, 50(6), 1083–1090. <https://doi.org/10.1037/0022-3514.50.6.1083>
- Frost, C., Czyzewska, M., & Kasch, L. (1992). *Reality Construction and Presentational Language: A Social-Cognitive Perspective on Prejudice*. *Jumanity and Society*, 16 (2).
- Gagliardone, I., Patel, A., & Pohjonen, M. (2014). *Mapping and Analyzing Hate Speech Online: Opportunities and challenges for Ethiopia*. *Electronic Journal*. doi:10.2139/ssrn.2601792
- Gagliardone, I., Pohjonen, M., Beyene, Z., Zerai, A., Aynekulu, G., Bekalu, M., ... Teferra, Z. (2016). *Mechachal: Online Debates and Elections in Ethiopia - From Hate Speech to Engagement in Social Media*. SSRN Electronic Journal . doi:10.2139/ssrn.2831369
- Gagliardone, I., Pohjonen, M., Zerai, A., Beyene, Z., Aynekulu, G., Gebrewolde, T., ... Moges, M. A. (2015). *Mechachal: Online Debates and Elections in Ethiopia. Report One: A Preliminary Assessment of Online Debates in Ethiopia*. *Electronic Journal*. doi:10.2139/ssrn.2782070
- Gallacher, J., Heerdink, M.W., & Hewstone, M. (2021). *Online Engagement between Opposing Political Protest Groups via Social Media is linked to Physical Violence of Offline Encounters*. *Social Media + Society*, 7(1).
- Galletta, A.M. (2013). *Mastering the Semi Structured Interview and Beyond*. NYU Press, New York. <https://doi.org/10.18574/nyu/9780814732939.001.0001>
- Gaylin, W. (2003). *Hatred: The Psychological Descent into Violence*
- Gebeyanu (March 5, 2022). የመሐል አገር ፖለቲከኞች ትግራይ እንድትገነጠል የሚፈልጉ ይመስለኛል. Accessed from <https://www.youtube.com/watch?v=ss9onfQcVNw>.

- Gelber, K., & McNamara, L. (2015). *The Effects of Civil Hate Speech Laws: Lessons from Australia*. *Law & Society Review*, 49(3), 631–664.
<http://www.jstor.org/stable/43670529>
- Gerbner, G. (1998). *Cultivation Analysis: an Overview*. *Mass Communication & Society*, 1, 3–4.
- Gessese, A.A. (2020). *Ethnic Nationalists Abuse of Media: Lessons of Yugoslavia and Rwanda for Ethiopia*. *European Scientific Journal*, 16(16), pp. 98–122.
- Giebels, E. (2008). *Categorization Processes and Conflict in Neighborhoods: the Role of Ingroup vs. Outgroup Categorization and Stereotypes*. IACM 21st Annual Conference Paper. <http://dx.doi.org/10.2139/ssrn.1298596>
- Giner-Sorolla, R., Leidner, B., & Castano, E. (2012). *Dehumanization, Demonization, and Morality shifting: Paths to moral certainty in extremist violence*. In Hogg M. A. & Blaylock D. L. (EDs.). *Extremism and the psychology of uncertainty*, 165–182, Wiley-Blackwell.
- Glaeser, E. L. (2002). *The Political Economy of Hatred*. Working Paper 9171 National Bureau of Economic Research Cambridge, MA 02138.
<http://www.nber.org/papers/w9171>
- Glaser, B.G. (2008). *Doing Quantitative Grounded Theory*. Mill Valley. Sociology Press
- Glaser, B. G. (2003). *The Grounded Theory Perspective II: Description's Remodeling of Grounded Theory Methodology*. Mill Valley, CA: Sociology Press.
- Glaser, B. G. (2001). *The Grounded Theory Perspective: Conceptualization Contrasted with Description*. Mill Valley, CA: Sociology Press.
- Glaser, B. (1998). *Doing Grounded Theory: Issues and Discussions*. Mill Valley, CA: Sociology Press.
- Glaser, B. G. (1978). *Theoretical Sensitivity: Advances in the Methodology of Grounded Theory*. Mill Valley: Sociology Press.
- Glaser, B. G. (1978). Theoretical sensitivity: Advances in the methodology of grounded theory. Mill
- Glaser, B. G. (1978). Theoretical sensitivity: Advances in the methodology of grounded theory. Mill

- Glaser, J., & Laudel, G. (2013). *Life With and Without Coding: Two Methods for Early-Stage Data Analysis in Qualitative Research Aiming at Causal Explanations*. *Qualitative Social Research*, 14(2); doi.org/10.17169/fqs-14.2.1886
- Glaser, B. G., and A. L. Strauss. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. New York, NY: Aldine de Gruyter.
- Global Peace Index (2023): *Measuring Peace in a Complex World*. Institute for Economics & Peace.
- Goffman, E. (1959). *The Presentation of Self in Everyday Life*. Garden City: Doubleday-Anchor
- Górka, M. (2019). Hate Speech – Causes and Consequences In Public Space. *Task Quarterly* vol. 23, No 2, 2019, pp. 233–244
- Graef, J., da Silva, R., & Lemay-Hebert, N. (2018). *Narrative, Political Violence and Social Change*. *Studies in Conflict & Terrorism*, 1–13. doi:10.1080/1057610x.2018.1452701
- Greenawalt, K. (1889) *Free Speech Justifications*, 89 COLUM. L. REV. 119 Available at: https://scholarship.law.columbia.edu/faculty_scholarship/84
- Groshek, J., & Groshek, M. C. (2013). *Agenda Trending: Reciprocity and the Predictive Capacity of Social Networking Sites in Intermedia Agenda Setting across Topics over Time*. *Media and Communication*, 1, 15–27. doi:10.12924/mac2013.01010015
- Guba, E. G., & Lincoln, Y. S. (1982). Epistemological and methodological bases of naturalistic inquiry. *Educational Communication & Technology Journal*, 30(4), 233–252.
- Guetterman, T. C., Babchuk, W. A., Howell Smith, M. C., & Stevens, J. (2017). *Contemporary Approaches to Mixed Methods–Grounded Theory Research: A Field-Based Analysis*. *Journal of Mixed Methods Research*, 13(2), 179-195.
- Habermas, T., & Bluck, S. (2000). *Getting a Life: The Emergence of the Life Story in Adolescence*. *Psychological Bulletin*, 126(5), 748–769.
- Hagos, N. S. (2018). Spreading Hatred: A study of Facebook in Ethiopia. *Global Media Review* 1 (2).
- Halaweh, M. (2018). *Integrating Social Media and Grounded Theory in a Research Methodology*. *Business Information Review*. doi:10.1177/0266382118809168
- Halaweh, M., Fidler, C., & McRobb, S. (2008). *Integrating the Grounded Theory Method and Case Study Research Methodology within IS Research: A Possible 'Road Map'*. *International Conference on Interaction Sciences*.

- Halperin, E., Canetti-Nisim, D., & Hirsch-Hoefler, S. (2009). *The Central Role of Group-Based Hatred as an Emotional Antecedent of Political Intolerance: Evidence from Israel*. *Political Psychology*, Vol. 30, No. 1, 2009
- Hardin, C. D., & Higgins, E. T. (1996). *Shared Reality: How social Verification Makes the Subjective Objective*. In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition*, Vol. 3. *The interpersonal context* (pp. 28–84).
- Hargittai, E. (2010). Digital na(t)ives? *Variation in internet skills and uses among members of the “net generation”*. *Sociol. Inq.* 80, 92–113.
- Harrington, E. (2004). *The Social Psychology of Hatred*. V(3)1. Paper presented at the Conference to Establish the Field of Hate Studies, Spokane, WA
- Haslam, N. (2006). *Dehumanization: An Integrative Review*. *Personality and Social Psychology Review*, 10, 252–264. doi.org/10.1207/s15327957pspr1003_4
- hate speech prevalence, severity, and natures, *Information, Communication & Society*,
- Heinze, E. (2016). *Hate Speech and Democratic Citizenship*. Oxford, UK: Oxford Univ. Press
- Hinton, A. L. (2005). *Why Did They Kill? Cambodia in the Shadow of Genocide*. University of California Press. Berkeley and Los Angeles
- Ho, S. S. (2016). *Spiral of Silence: The International Encyclopedia of Political Communication*, 1–5. doi:10.1002/9781118541555.wbiepc0
- Hodson, G., Kteily, N.S., & Hoffarth, M.R. (2014). *Of filthy pigs and subhuman mongrels: Dehumanization, disgust, and intergroup prejudice*.
- Hogg, M. & Smith, J. (2007). *Attitudes in Social Context: A social identity perspective*. *European Review of Social Psychology - Eur Rev Soc Psychol.* 18. 89-131. 10.1080/10463280701592070.
- Howard, J. (2019). *Free Speech and Hate Speech*. *Annual Review of Political Science*. Vol. 22: pp. 93-109 <https://doi.org/10.1146/annurev-polisci-051517-012343>
- IFES, (2018 b). *Social Media, Disinformation and Electoral Integrity: Working Paper*. https://www.ifes.org/sites/default/files/ifes_working_paper_social_media_disinformation_and_electoral_integrity_august_2019.pdf
- Ilibagiza, I. & Erwin, S. (2006). *Left to Tell. Discovering God Amidst the Rwandan Holocaust*. Carlsbad, Hay House.

iMAGE eTHIOPIS (2018). *Ethiopian Politics – Dr abiy ahmed speech about meskel adebabay Grenade attack cause 2018* [Video]. Accessed from <https://www.youtube.com/watch?v=sK8zdpJTh1Y>

Institute for Economics & Peace. Global Peace Index (2023): Measuring Peace in a Complex World, Sydney, June 2023. Retrieved on 10/30/2023 from <https://www.visionofhumanity.org/wp-content/uploads/2023/06/GPI-2023-Web.pdf>

International Foundation for Electoral Systems, IFES (2018 a). *Disinformation Campaigns and Hate Speech: Exploring the Relationship and Programming Interventions*. https://www.ifes.org/sites/default/files/2019_ifes_disinformation_campaigns_and_hate_speech_briefing_paper.pdf

Jääskeläinen, T. (2020). *Countering Hate Speech Through Arts and Arts Education: Addressing intersections and policy implications*. Policy Futures in Education, 18(3), 344-357. doi.org/10.1177/1478210319848953

Jackson, R. L., & Hogg, M. L. (2010). *Encyclopedia of Identity* (Vol. 1). California. Sage.

Jakubowicz, A., Dunn, K., Mason, G., Paradies, Y., Bliuc, A., Bahfen, N., Oboler, A. Atie R. Connelly, R. (2017). *Cyber Racism and Community Resilience: Strategies for Combating Online Race Hate*. Palgrave Macmillan

James, William (1907). *Pragmatism's Conception of Truth. Lecture 6 in Pragmatism: A new name for some old ways of thinking*. New York: Longman Green and Co 76-91.

Johnston, K., Tanner, M., Lalla, N., & Kawalski, D. (2013). *Social capital: the benefit of Facebook 'friends'*. *Behaviour & Information Technology*, 32(1), 24-36.

Joinson, A.N. (2001). *Self-disclosure in Computer-mediated Communication: The role of self-awareness and visual anonymity*. *Eur J Soc Psychol*, 31: 177-192

Jost, J., Ledgerwood, A., & Hardin, C. (2007). *Shared Reality, System Justification, and the Relational Basis of Ideological Beliefs*. *Social and Personality Psychology Compass* 10.1111/j.1751-9004.2007.00056.x.

Justitia (2020). Global Handbook on Hate Speech Laws. from https://futurefreespeech.com/wp-content/uploads/2020/11/Report_Global-Handbook-on-Hate-Speech-Laws.pdf

Kang, H., Keunmin B., Shaoke Z., & Shyam S. (2011). *Source Cues in Online News: Is The Proximate Source More Powerful than Distal Sources?* *Journalism & Mass Communication Quarterly* 88 (4): 719–736. doi:10.1177/107769901108800403

- Kaplan, A. M. & Haenlein, M. (2010). *Users of the World, Unite! The challenges and opportunities of social media*. *Business Horizons*, 53(1), 59–68.
- Katz, E., Blumler, J. G., & Gurevitch, M. (1974). *Uses and Gratifications Research*. *Public Opinion Quarterly*, 37(4), 509–523. <http://www.jstor.org/stable/2747854>
- Keen, S. (1991). *Faces of the Enemy: Reflections of the Hostile Imagination*. Harper & Row.
- Kennedy, B. G., Trager J, Atari M, Hoover J, Davani, M., & Dehghani, M. (2023). *The (moral) Language of Hate*. *PNAS Nexus*. 11;2(7). doi: 10.1093/pnasnexus/pgad210.
- Kim, J. Y. (2018). *A study of Social Media Users' Perceptual Typologies and Relationships to Self-identity and Personality*. *Internet Res.* 28, 767–784. doi: 10.1108/IntR-05-2017-0194
- Kinfe, M.Y. (2017). *Fake News' and Its Discontent in Ethiopia – a reflection*. *Mekelle University Law Journal*, 5(1).
- Kokil, J., Alvin Z., Yphtach L., Jana E., & Sophie L. (2022). *Beyond Anonymity: Network Affordances, Under Deindividuation, Improve Social Media Discussion Quality*. *Journal of Computer-Mediated Communication*, Volume 27, Issue 1. doi.org/10.1093/jcmc/zmab019
- Krämer, N. C., & Winter, S. (2008). *Impression Management 2.0: The Relationship of Self-Esteem, Extraversion, Self-efficacy, and Self-presentation within Social Networking Sites*. *Journal of Media Psychology: Theories, Methods, and Applications*, 20(3), 106–116. <https://doi.org/10.1027/1864-1105.20.3.106>
- Kressel, N. J. (2002). *Mass Hate: the global rise of genocide and terror*. New York. Westview.
- Krithika, G. K., & Kumar, K. M. (2018). *The Social Media user: a theoretical background to the development of social media user typology*. *ELK Asia Pac. J. Mark. Retail Manag.* 9. doi: 10.31511/EAPJMRM.2018v09i04001
- Kteily, N. S., & Bruneau, E. (2017). *Darker Demons of Our Nature: The Need to (Re)Focus Attention on Blatant Forms of Dehumanization*. *Current Directions in Psychological Science*, 26(6), 487-494.
- Kteily, N., Hodson, G., & Bruneau, E. (2016). They see us as less than human: Metadehumanization predicts intergroup conflict via reciprocal dehumanization. *Journal of Personality and Social Psychology*, 110 (3), 343–370.

- Lai, L.S., & To, W. (2015). *Content Analysis of Social Media: A Grounded Theory Approach*. *Journal of Electronic Commerce Research*, 16, 138.
- Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. University of Chicago, Chicago, IL.
- Lampropoulos, G., Anastasiadis, T., Siakas, K., & Siakas, E. (2022). *The Impact of Personality Traits on Social Media Use and Engagement: An overview*. *International Journal on Social and Education Sciences*, 4(1), 34-51.
- Lasser, J., Herderich, A., Garland, J., Aroyehun, S. T., Garcia, D., & Galesic, M. (2023). *Collective Moderation of Hate, Toxicity, and Extremity in Online Discussions*. *arXiv*. doi.org/10.48550/arXiv.2303.00357
- Lee, R (2021). *Myanmar's Rohingya Genocide: Identity, History and Hate Speech*. I.B. Tauris. London.
- Lee, J., & Thorson, E. (2008). *The Impact of Celebrity–Product Incongruence on the Effectiveness of Product Endorsement*. *Journal of Advertising Research*, 48, 433 - 449.
- Voorveld, H. A. M. (2019). *Brand Communication in Social Media: A Research Agenda*. *Journal of Advertising*, 48(1), 14–26.
<https://doi.org/10.1080/00913367.2019.1588808>
- Lerner, M. J. (1980). *The Belief in a Just World: A Fundamental Delusion*, DOI: 10.1007/978-1-4899-0448-5_2
- Leuven, K., Pandelaere, M., Vanbeselaere, N., & Dewitte, S. (2002). *Politics and Basking-in-Reflected-Glory: A Field Study in Flanders*. *Basic and Applied Social Psychology*, 24(3), 205–214, doi: 10.1207/153248302760179129
- Levin, J. & McDevitt, J. (1993). *Hate Crime: the Rising Tide of Bigotry and Bloodshed*. New York. Springer
- Levine, J. M & Hogg, M. A. (EDs.) (2010). *Encyclopedia of Group Processes and Intergroup Relation*. London. Sage.
- Levitt H. M., Motulsky, S.L., Wertz, J. F., Morrow S.L., Ponterotto, J. G (2017). *Recommendations for Designing and Reviewing Qualitative Research in Psychology: Promoting Methodological Integrity*. *Qualitative Psychology*, 4 (1)
- Leyens J., Demoulin S., Vaes J., Gaunt R., & Paladino M. P. (2007). *Infra-humanization: The wall of group differences*. *Social Issues and Policy Review*, 1, 139–172.

- Li, Y. & Hills, T. (2021). *Language Patterns of Outgroup Prejudice*. *Cognition*, doi.org/10.1016/j.cognition.2021.104813
- Lilo Media ሲኤ ሚዲያ (July 25, 2021). *Agegnehu Teshager Genocidal Speech against Tigray* (Video file). Accessed from <https://www.youtube.com/watch?v=EcQCFIzDAho>
- Lincoln, Y. S. & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage
- Livesley, W. J. (2001). (Ed). *Handbook of Personality Disorders: Theory, Research, and Treatment*. New York: Guilford.
- Livingston, M. M. & Pace, L. A. (2005). *Protecting Human Subjects in Internet Research*. *EJBO Electronic Journal of Business Ethics and Organization Studies*;10(1), 35-41
- Livingstone, A. G., Spears, R., Manstead, A. S. R., Bruder, M., & Shepherd, L. (2011). We feel, therefore we are: Emotion as a basis for self-categorization and social action. *Emotion*, 11(4), 754–767. doi.org/10.1037/a0023223
- Lord, C. G., Ross, L., & Lepper, M.R. (1979). *Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence*. *Journal of Personality and Social Psychology*, 37 (11)
- Luhtanen, R. & Crocker, J. (1991). *Collective self-esteem scale: Self-Evaluation of One's Social Identity*. *Personality and Social psychology bulletin*; 18(3), 302-318
- Macrae, C. N., Bodenhausen, G. V., & Milne, A. B. (1995). *The Dissection of Selection in Person Perception: Inhibitory Processes in Social Stereotyping*. *Journal of Personality and Social Psychology*, 69(3), 397–407.
- Maddrell, P. (2015). *The Image of the Enemy: Intelligence Analysis of Adversaries since 1945*. Georgetown University Press
- Mahl, D & Schäfer, S. M. (2023). *Conspiracy Theories in Online Environments: An interdisciplinary literature review and agenda for future research*. *new media & society*, 25(7), 1781–1801. doi.org/10.1177/14614448221075759
- Maior, G. R., Olson, J. M., & Bush, J. E. (1997). *Telling Jokes that Disparage Social Groups: Effects on the joke teller's stereotypes*. *Journal of Applied Social Psychology*, 27, 1986–2000. doi .org/10.1111/j.1559-1816.1997.tb01636.x
- Malešević, S. (2004). *The Sociology of Ethnicity*. London. Sage.
- Marcus, K. L. (2012). *Accusation in a Mirror*. *Loyola University Chicago Law Journal*, 43 (2),

- Markham, A. & Buchanan, E. (2012). *Ethical Decision--Making and Internet Research: Recommendations from the AoIR Ethics Working Committee (Version 2.0)* Available at:<http://aoir.org/reports/ethics2.pdf>
- Martínez, C. A., van Prooijen, J.-W., & Van Lange, P. A. M. (2022). *A threat-based Hate Model: How symbolic and realistic threats underlie hate and aggression*. *Journal of Experimental Social Psychology*, 103, 1–13. doi.org/10.1016/j.jesp.2022.104393
- Matamoros-Fernández, A & Farkas, J. (2021). *Racism, Hate Speech, and Social Media: A Systematic Review and Critique*. *Television & New Media* 2021, Vol. 22(2) 205–224.
- Mathew, B., Dutt, R., Goyal, P., & Mukherjee, A. (2018). *Spread of Hate Speech in Online Social Media*. *Proceedings of the 10th ACM Conference on Web Science - WebSci'19*. doi:10.1145/3292522.3326034
- Megersa, M. & Sinishaw, S. (2019). *Unveiling Ethnic Hate Speech on Social Media: A Social Psychological Examination*. *Ethiopian journal of Behavioral Studies*. V 2(1)
- Mercy Corps (2021). *Social Media and Conflict: Understanding Risks and Resilience Research Summary and Policy Brief*.
<https://www.mercycorps.org/sites/default/files/2021-08/Digital-Conflict-Research-Summary-and-Policy-Brief-073021.pdf>
- Mereja TV (April 27, 2021). *የእኛ አለ-ላ ሰለሞን እብደት* (Video file). Accessed from <https://www.facebook.com/watch/?ref=saved&v=1179375379178905>
- Meta (2021). *Facebook Community Standard*. Accessed from https://www.facebook.com/communitystandards/hate_speech
- Meta (2022). *Longstanding Work to Protect People in Ethiopia*.
<https://about.fb.com/news/2021/11/update-on-ethiopia/>
- Meta (2023). *Hate Speech Policy Details*: <https://transparency.fb.com/policies/community-standards/hate-speech/>
- Mirbabaie, M., & Zapatka, E. (2017). *Sensemaking in Social Media Crisis Communication : a case study on the Brussels bombings in 2016*. In *Proceedings of the 25th European Conference on Information Systems*; 2169–2186.
- Miškolci, J., Lucia K., & Edita R. (2018). *Countering Hate Speech on Facebook*. *Social Science Computer Review*. doi.org/10.1177/0894439318791786
- Mkono, M. (2015). *Troll Alert: Provocation and Harassment in Tourism and Hospitality in Social Media*. *Curr. Issues Tour*. 21, 791–804. doi: 10.1080/13683500.2015.1106447

- Mondal, M., Silva, L. A., Correa, D., & Benevenuto, F. (2018). *Characterizing usage of explicit hate expressions in social media*. *New Review of Hypermedia and Multimedia*, 24(2), 110–130. doi:10.1080/13614568.2018.1489001
- Moore, J., Magee, S., Gamrekidze, E., & Kowalewski, J. (2019). *Social Media Mourning: Using Grounded Theory to Explore How People Grieve on Social Networking Sites*. *Journal of Death and Dying*, 79(3), 231-259.
<https://doi.org/10.1177/0030222817709691>
- Morrock, R. (2010). *The Psychology of Genocide and Violent Oppression: A Study of Mass Cruelty from Nazi Germany to Rwanda*. McFarland & Company. London.
- Muluken, A.C., Mulatu A.M., & Biset Ayalew (2021). *Social Media Hate Speech in the Walk of Ethiopian Political Reform: Analysis of Hate Speech Prevalence, Severity and Nature*. *Journal of Information, Communication & Society*, (26) 1. 1-20.
- Muluken, Tesfaw (July 16, 2021). "ወላይታን እንደሰው ቆጠርናቸው፤ አሮሞም እንደሰው የለውጥናቸው እኛ ነን" ይላሉ የብሄር ብሄረሰብ ጠበቃ ነኝ የሚሉት (Video file). Accessed from
<https://www.facebook.com/watch/?ref=saved&v=403268334442223>).
- Myers, D. (2010). *Social Psychology* (10thed). New York: McGraw-Hill
- Nadim, M., & Fladmoe, A. (2016). *Hate Speech, report1: Research on the nature and extent of hate speech*. *Norwegian Institute for Social Research*. Retrieved from www.samfunnsforskning.no
- Nadler, A., Malloy, T. E., & Fisher J. D. (2008). *The Social Psychology of Intergroup Reconciliation*. Oxford University Press. New York.
- Navarro, J. (2013). *The Psychology of Hatred*. *The Open Criminology Journal*, 6, 10-17.
- Noel, J. G., Wann, D. L., & Branscombe, N. R. (1995). *Peripheral Ingroup Membership Status and Public Negativity toward Outgroups*. *Journal of Personality and Social Psychology*, 68(1), 127–137. doi.org/10.1037/0022-3514.68.1.127
- Nigatu (2021): Social media hate speech in the walk of Ethiopian political reform: analysis
 oNoel, J. G., Wann, D. L., & Branscombe, N. R. (1995). Peripheral ingroup membership status and public negativity toward outgroups. *Journal of Personality and Social Psychology*, 68(1), 127–137. <https://doi.org/10.1037/0022-3514.68.1.127>
- Noelle-Neumann, E. (1993). *The Spiral of Silence: Public opinion, our social skin* (2nd ed.). Chicago, IL: University of Chicago Press.

- O'Donovan, T., Fournelle, C., Brdiczka, O., Shen, J., Liu, J., & Moore, E. (2013). Characterizing Users' Behavior and Information Propagation on a Social Multimedia Network. IEEE International Conference on Multimedia and Expo Workshops (ICMEW). doi:10.1109/icmew.2013.6618395
- Opotow, S. & McClelland, S. I. (2007). *The Intensification of Hating: A Theory Social Justice Research*, Vol (20) 1; DOI: 10.1007/s11211-007-0033-0
- Oppenheimer, L. (2006). *The Development of Enemy Images: A Theoretical Contribution*. Journal of Peace Psychology, 12(3), 269–292.
- Ortiz, J., Young, A. Michael D. M. Bedeley, R.T., & Carbaugh, D. (2019). *Giving Voice to the Voiceless: The Use of Digital Technologies by Marginalized Groups*. Accessed from <https://core.ac.uk/download/pdf/301380156.pdf>
- Osborne, R. E & Frost, C. J (2004). *The Anatomy of Hatred: Multiple Pathways to the Construction of Human. Humanity & Society*. 28(5)
- Ostapenko, I.V. & Sosniuk O (2019). *Psychological Features of Social Meida Users*. Ukrainian Psychological journal. doi.org/10.17721/upj.2019.2(12).12
- Özgüven, N., & Mucan, B. (2013). *The Relationship between Personality Traits and Social Media Use*. Social Behavior and Personality: An International Journal, 41(3), 517–528. doi:10.2224/sbp.2013.41.3.517
- Pariser, E. (2011). *The Filter Bubble: What The Internet Is Hiding from you?* New York: Penguin Press
- Parmelee, J. H. (2013). *The Agenda-building Function of Political Tweets*. New Media & Society, 16, 434–450. doi:10.1177/1461444813487955
- Pace, L. & Livingston, M. (2005). Protecting Human Subjects in Internet Research. Electronic Journal of Business Ethics and Organization Studies 10(1): 35-41
- Pennington, D.C. (2000). *Social Cognition*. Routledge. London
- Pennycook, G., & Rand, D. G. (2019). *Fighting Misinformation on Social Media using crowdsourced judgments of news source quality*. Proceedings of the National Academy of Sciences, 116, 2521–2526.
- Pettigrew, T. F., & Meertens, R. W. (1995). *Subtle and Blatant Prejudice in Western Europe*. European Journal of Social Psychology, 25(1), 57–75.

- Pew Research Center (2014). *Online Harassment*. Washington, D.C.. Available at <https://www.pewresearch.org/internet/2014/10/22/online-harassment/>
- Pew Research Center (2021). *The State of Online Harassment*. Available at <https://www.pewresearch.org/internet/2021/01/13/the-state-of-online-harassment/>
- Pohjonen, M. (2018). *Horizons of Hate: A Comparative approach to Social Media Hate Speech*. VOX-Pol Network of Excellence.
- Post, R.(2009). *Hate Speech*. In Ivan Hare and James Weinstein (eds.), *Extreme Speech and Democracy*. Oxford University Press.
- Postmes, T., & Spears, R. (1998). *Deindividuation and Anti-normative Behavior: A meta-analysis*. *Psychological Bulletin*, 123, 238–259.
- Pousti, M., Urquhart, C., Burstein, F., & Linger, H. (2013). *Methodological implications of Social Media as a research setting for IS healthcare studies: Reflections from a grounded theory study*. 24th Australasian Conference on Information Systems. 4-6 Dec 2013
- Pratto, F., & Glasford, D. E. (2008). *Ethnocentrism and the Value of a Human Life*. *Journal of Personality and Social Psychology*, 95(6), 1411–1428. <https://doi.org/10.1037/a0012636>
- Pronin, E., Kruger, J., Savtisky, K., & Ross, L. (2001). *You Don't Know Me, But I Know You: The illusion of asymmetric insight*. *Journal of Personality and Social Psychology*, 81(4), 639–656. <https://doi.org/10.1037/0022-3514.81.4.639>
- Prosperity Party - ብልፅግና (December 20, 2021). *ወያኔን መረዳት ሳይሆን ማጥፋት ነው የሚቻለው* (Video file). from <https://www.facebook.com/watch/?ref=saved&v=986232581984357>.
- Reed, A., & Aquino, K. (2003). *Moral Identity and the Expanding Circle of Moral Regard Toward Out-groups*. *Journal of personality and social psychology*, 84 6, 1270-86
- Rempel, J. K & Burriss C.T (2005). *Let Me Count the Ways: An integrative theory of love and hate*. *Journal of Personal Relationship*, (12) 2 . P. 297-313.
- Ribeiro, M. H., Calais, H.P., Santos, Y. A., Almeida V.,& Meira W. (2018). *Characterizing and Detecting Hateful Users on Twitter*. *International Conference on Web and Social Media*. doi.org/10.1609/icwsm.v12i1.15057
- Richards, H. K., Baele, S.J., & Coan, T. (2019). *Studying “Radio Machete”: Towards a Robust Research Programme*. *Journal of Genocide Research*, 21:4, 525-539, DOI: 10.1080/14623528.2019.1652017

- Rieber, R. (1991) (Ed.). *The Psychology of War and Peace: The Image of the Enemy*. Springer
- Rimé, B. & Christophe, V. (1997). *How Individual Emotional Episodes Feed Collective Memory*. In Pennebaker, J.W., Paez, D., Rim, B., Paez, D. (1997) (Eds.). *Collective Memory of Political Events: Social Psychological Perspectives*. Routledge.
- Rogers, R. W., & Prentice-Dunn, S. (1981). *Deindividuation and Anger-mediated Interracial Aggression: Unmasking Regressive Racism*. *Journal of Personality and Social Psychology*, 41(1), 63.
- Romero, D., Galuba, W., Asur, S. & Huberman, B. (2011). *Influence and Passivity in Social Media*. *Inf. Syst. J.* 1-9. 10.1145/1963192.1963250.
- Ross, M. H. (2001). *Psychocultural Interpretations and Dramas: Identity Dynamics in Ethnic Conflict*. *Political Psychology*, 22(1), 157–178. doi:10.1111/0162-895x.00231
- Roussos, G., & Dovidio, J. F. (2018). *Hate Speech Is in the Eye of the Beholder: The Influence of Racial Attitudes and Freedom of Speech Beliefs on Perceptions of Racially Motivated Threats of Violence*. *Social Psychological and Personality Science*, 9(2), 176-185. <https://doi.org/10.1177/1948550617748728>
- Royzman, E., McCauley, R., & Rozin, P. (2004). *From Plato to Utman: Four ways to think about Hate*. <https://cpb-us-w2.wpmucdn.com/web.sas.upenn.edu/dist/7/206/files/2016/09/HateRoyzman-Sternberg03-1uadz8i.pdf>
- Ruse, M. (2022). *Why We Hate: Understanding the Roots of Human Conflict*. New York. Oxford University Press
- Ryan, T., Chester, A., Reece, J., & Xenos, S. (2014). *The Uses and Abuses of Facebook: A Review of Facebook Addiction*. *Journal of behavioral addictions*, 3(3), 133-148.
- Sandelowski, M. (2000). Focus on Research Methods: Whatever happened to qualitative description? *Research in Nursing and Health*, 23(4), 334-340
- Saucier, D. A., O'Dea, C. J., & Strain, M. L. (2016). The Bad, the Good, the Misunderstood: The Social Effects of Racial Humor. *Translational Issues in Psychological Science*, 2(1), 75–85. <https://doi.org/10.1037/tps0000059>
- Schieb, C. & Preuss, M. (2016). *Governing Hate Speech by Means of Counterspeech on Facebook*. 66th ICA Annual Conference, at Fukuoka, Japan, pp. 1-23.
- Schreier, M. (2012). *Qualitative Content Analysis in Practice*. Thousand Oaks, CA: Sage

- Seidman, G. (2020). *Personality Traits and Social Media Use*.
doi.org/10.1002/9781119011071.iemp0295
- Sellars, A. F. (2016). *Defining Hate Speech*. Berkman Klein Center Research Publication
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2882244
- Shachaf, P., & Hara, N. (2010). *Beyond Vandalism: Wikipedia Trolls*. *Journal of Information Science*, 36(3), 357-370. <https://doi.org/10.1177/0165551510365390>
- Banaji, S. & Bhat, R. (2022). *Social Media and Hate*. Routledge. London
- Sherif, M. (1966). *Group Conflict and co-operation: Their Social Psychology*. London: Routledge
- Shirayev, E.B., & Levy, D.A. (2013). *Cross-cultural Psychology: Critical Thinking and Contemporary Application* (5th Ed.). Boston: Allyn & Bacon
- Shumete, Gizaw (2020). አግጥሞ (video file). Accessed at January 29 2020. from
<https://www.facebook.com/shumete.woldeamanuel/videos/10156696594256615/>
- Sidanius, J., & Pratto, F. (1999). *Social Dominance Theory*. *Handbook of theories of social psychology*
- Siegel, A. A., & Badaan, V. (2020). #No2Sectarianism: Experimental Approaches to Reducing Sectarian Hate Speech Online. *American Political Science Review*, 1–19. doi:10.1017/s0003055420000283
- Silverman, D. (2014). *Interpreting Qualitative Data* (5thed.). Los Angeles. Sage.
- Silverstein, B. (1987). *Toward a Science of Propaganda*. *Political Psychology*, 8(1), 49–59.
<https://doi.org/10.2307/3790986>
- Simpson, R. (2013). *Dignity, Harm, and Hate speech*. *Law Philos.* 32(6):701–28
- Singer, P (1981). *Expanding the Circle of Moral Concern*. Princeton University Press
- Singer P. W & Brooking E T. (2018). *Like War: Weaponizing Social Media*. Houghton Mifflin Harcourt. New York
- Skelly, J. M. (2022). *War on Hate: How to Stop Genocide, Fight Terrorism, and Defend Freedom: Israel Journal of Foreign Affairs*, 16(2), 271–275.
- Skitka, L. J. (2002). *Do the Means Always Justify the Ends, or Do the Ends Sometimes Justify the Means? A value model of justice reasoning*. *Personality and Social Psychology Bulletin*, 28(5), 588–597
- Skjerdal, T. & Mulatu A.M. (2021). *The Ethnicification of the Ethiopian media: A research report*. Addis Ababa. Available at: <https://ssrn.com/abstract=378541>

- Snijders, R. & Helms, R. (2014). *Analyzing Social Influence through Social Media, A Structured Literature Review*. Proceedings of the 7th IADIS International Conference on Information Systems, 28 February – 2 March, Madrid.
- Solomon, E. (1977). *Brand Hate*. In Fischer, A., Halperin, E., Canetti, D., & Jasini, A. (2018). Why We Hate. *Emotion Review*, 10 (4); 309–320
- Soral, W., Bilewicz, M., & Winiewski, M. (2018). *Exposure to Hate Speech Increases Prejudice through Desensitization*. *Journal of Aggressive behavior*, 44(2): 109-220
- Spinda, John. (2011). *The Development of Basking in Reflected Glory (BIRG) and Cutting off Reflected Failure (CORF) Measures*. *Journal of Sport Behavior*; 34. 392-420.
- Sripokangkul, S., Crumpton, C. D., & Draper, J. (2020). *Demonization and Violence as Tools of Institutional Power: the social cage of Thailand*. *Social Identities*, 26(4), 446–460. doi.org/10.1080/13504630.2020.1767056
- Stanton, G.H (1998). *Ten Stages of Genocide*. The Genocide Education Project. accede from www.genocideeducation.org/wp-content/uploads/2016/03/ten_stages_of_genocide.pdf
- Statista (2023). <https://www.statista.com/statistics/1312554/social-media-users-by-platform-in-ethiopia/>
- Staub, E. (1992). *The Roots of Evil: The origins of genocide and other group violence*. Cambridge: Cambridge University Press.
- Staub, E. (2005). *The Origins and Evolution of Hate, with Notes on Prevention*. In *Psychology of Hate* (pp. 51–65). Doi.org/10.1037/10930-003
- Stein, R. M. (2016). *The Face of the Enemy: Images, Stereotypes, and Perception of Threat*. George Washington University.
- Steinfeld, C., Ellison, N. B., & Lampe, C. (2008). *Social Capital, Self-esteem, and Use of Online Social Network Sites: A longitudinal analysis*. *Journal of applied developmental psychology*, 29(6), 434-445.
- Sternberg, R. J. (2003). *A Duplex Theory of Hate: Development and Application to Terrorism, Massacres, and Genocide*. *Review of General Psychology*, 7(3), 299–328. doi:10.1037/1089-2680.7.3.299
- Sternberg, R. J. (2005). *The Psychology of Hate*. Washington: APA.
- Sternberg, R. J. (2018). *FLOTSAM: A model for the development and transmission of hate*. *Journal of Theoretical Social Psychology*. doi:10.1002/jts5.25

- Sternberg, R. J. (Ed.). (2020). *Perspectives on Hate: How It Originates, Develops, Manifests, and Spreads*. APA. Available online from: <http://www.jstor.org/stable/j.ctv1chrvt9>
- Sternberg, R. J. & Sternberg, K. (2008). *The Nature of Hate*. Cambridge University Press
- Strate, L. (2008). *Studying Media as Media: McLuhan and the Media Ecology Approach*. MediaTropes eJournal Vol I (2008): 127–142
- Strate, L. (2017). *Media Ecology*. An approach to understanding the human condition. Peter Lang.
- Strauss, A., & J. Corbin (1998). *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. 2nd ed. Thousand Oaks, CA: Sage
- Suddaby, R. (2006). *From the Editors: What Grounded Theory is Not*. Academy of Management Journal, 49(4), 633–642. doi:10.5465/amj.2006.22083020
- Sunstein, C. (1993). *Democracy and the Problem of Free Speech*. New York: Free Press.
- Sunstein, C. (2009). *Going to Extremes: How Like-Minds Unite and Divide*. New York Oxford.
- Surma, J. (2016). *Social Exchange in Online Social Networks. The reciprocity phenomenon on Facebook*. *Computer Communications*, 73, 342–346. doi:10.1016/j.comcom.2015.06.017
- Swann, W. B., Gómez, A., Seyle, D. C., Morales, J., & Huici, C. (2009). *Identity Fusion: the interplay of personal and social identities in extreme group behavior*. *Journal of personality and social psychology*, 96(5), 995.
- Tajfel, H. (1982). *Social Psychology of Intergroup Relations*. Ann. Re., Psychol. 1982. 33:1-39
- Tajfel, H., & Turner, J. C. (1979). *An integrative Theory of Intergroup Conflict*. In W. G. Austin, & S. Worchel (Eds.), *The Social Psychology of Intergroup Relations* (pp. 33-37). Monterey, CA: Brooks/Cole.
- Tamirat, Strong, (December 21, 2021). ተናዳላች አይገልፀውም እርር ከማለቷ የተነሳ ... (Video). Accessed from <https://www.facebook.com/watch/?ref=saved&v=423859649440777>).
- Tamiru, A. (September 2, 2020). ባሳለፍነው እሁድ ገዳ በሚባለው የወረራ፣ የጦርነትና የዘር ማጥፋት ሥርዓትና ... (Video file). Retrieved from <https://www.facebook.com/achamyelah.tamiru.3/videos/3397427523612506>
- Tandoc, E., Zheng, L., & Rich, L. (2017). *Defining Fake News: A typology of scholarly definitions*. Digital Journalism.1-17. 10.1080/21670811.2017.1360143.

- Tewodros, W. (2019). *Ethiopia's Hate Speech Predicament: Seeking Antidotes beyond a Legislative Response*, *African Journalism Studies*, 40:3, 123-139, DOI: 10.1080/23743670.2020.1729832
- TghatMedia (3 January 2022). *Weyane is a Cancer* [video file]
<https://twitter.com/TghatMedia/status/1477807860855840770>
- The Guardian (2022). Facebook Let Vigilantes in Ethiopia Incite Ethnic Killing'.
<https://www.theguardian.com/technology/2022/feb/20/facebook-lets-vigilantes-in-ethiopia-incite-ethnic-killing>
- Tiemessen, A. (2004). After Arusha, Gacaca Justice in Post-genocide Rwanda. *African Studies Quarterly*, 8(1), 1.
- Timmermann, Wibke (2008). *Counteracting Hate Speech as a Way of Preventing Genocidal Violence: Genocide Studies and Prevention: An International Journal: Vol. (3) 3*
- Tipler, C. & Ruscher, J. B (2014). *Agency's Role in Dehumanization: Non-human Metaphors of Out-groups*. *Social and Personality Psychology Compass* (8)5; 214–228,
- Turner, J.C. (1985). *Social Categorization and the Self-Concept: A Social Cognitive Theory of Group Behavior*. In: Lawler, E., Ed., *Advances in Group Processes*, JAI Press, 77-121.
- Turner, J. C. & Oakes, P.J.(1986). *The Significance of the Social Identity Concept for Social Psychology with Reference to Individualism, Interactionism and Social influence*. *British Journal of Social Psychology*, 25: 237-252. doi:10.1111/j.2044-8309.1986.tb00732.x
- Twitter (2016). *Hateful Conduct Policy*. <https://goo.gl/NxR4sR>.
- Tylor, K. (2004). *Brainwashing: the science of thought control*. Oxford. New York
- UN (2015). *Transforming our world: The 2030 Agenda for Sustainable Development*.
<https://sdgs.un.org/sites/default/files/publications/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf>
- UN (2020). *United Nations Strategy and Plan of Action on Hate Speech: Detailed Guidance on Implementation for United Nations Field Presences*.
- UNESCO (2015). *Countering Online Hate Speech*. Author. Accessible at
<file:///C:/Users/wer/Downloads/233231eng.pdf>

- UNESCO (2023). Addressing Hate Speech through Education: A guide for policy-makers. Accessible at https://www.un.org/en/genocideprevention/documents/publications-and-resources/Addressing_hate_speech_through_education_A_guide_for_policy_makers.pdf
- Urquhart, C. (2013). *Grounded theory for Qualitative Research*. SAGE Publications, Ltd, <https://doi.org/10.4135/9781526402196>
Valley: Sociology Press.
Valley: Sociology Press.
- Verkuyten, M., & Brug, P. (2004). Multiculturalism and group status: The role of ethnic identification, group essentialism and protestant ethic. *European Journal of Social Psychology*, 34(6), 647–661. <https://doi.org/10.1002/ejsp.222>
- Volkan, V. D. (2001). *Transgenerational Transmissions and Chosen Traumas: An Aspect of Large-Group Identity*. *Group Analysis*, 34(1), 79-97.
- Waldron, J.(2012). *The Harm in Hate Speech*. Cambridge: Harvard University Press
- Walther, J.B. (2022). *Social Media and Online Hate*. *Current Opinion in Psychology*. V (45) doi.org/10.1016/j.copsyc.2021.12.010
- Wasilewski, K (2019). *Hate speech and identity politics. An intercultural communication perspective*. *Journal of Przegląd Europejski*; (3); P. 175-187
- Watt, D. (2007). *On Becoming a Qualitative Researcher: The Value of Reflexivity*. *The Qualitative Report* Volume 12 Number 1 March 2007 82-101
- Westermann, E. B. (2021). *Drunk on Genocide Alcohol and Mass Murder in Nazi Germany*. Cornell University Press. Ithaca and London
- Wibke, T. (2008). *Counteracting Hate Speech as a way of Preventing Genocidal Violence*: *International Journal of Genocide Studies and Prevention*: Vol. 3: Iss 3: Article 8. Retrieved from <http://scholarcommons.usf.edu/gsp/vol3/iss3/8>
- Winiewski, M., Soral, W., Świdarska, A., Bulska, D. (2017). *Contempt Speech, Hate Speech: research on Verbal Violence against Minority groups Warsaw*: Accessed from https://www.ngofund.org.pl/wpcontent/uploads/2017/02/Contempt_Speech_Hate_Speech_Full_Report.pdf
- Woolley, S. & Howard P. (2018). *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media*. Oxford University Press. New York
- World Justice Project (WJP) (2023). Rule of Law Index. Retrived on 10/30/2023 from <https://worldjusticeproject.org/rule-of-law-index/downloads/WJPIndex2023.pdf>

- Yanal, Z. (2022). *The Ideology of Hatred: The Psychic Power of Discourse*. Fordham University Press. New York.
- Yardi, S., & Boyd, D. (2010). *Dynamic Debates: An Analysis of Group Polarization over Time on Twitter*. *Bulletin of Science, Technology & Society*, 30(5), 316-327. <https://doi.org/10.1177/0270467610380011>
- YouTube (2021). YouTube Policy. <https://support.google.com/youtube/answer/2801939?hl=en#>
- Young, C. (2008). *Explaining the Conflict Potential of Ethnicity*. In Darby and Ginty (2008) (Eds). *Contemporary Peacemaking Conflict, Peace Processes and Post-war Reconstruction*. Palgrave Macmillan. London.
- Yzerbyt, V., & Rogier, A. (2001). Blame it on the group: Entitativity, subjective essentialism, and social attribution. In J. T. Jost & B. Major (Eds.), *The psychology of legitimacy: Emerging perspectives on ideology, justice, and intergroup relations* (pp. 103–134). Cambridge University Press.
- Zahidi, Saadia. (2023, January 12). *Solving Global Crises Requires A Return To Basics*. Time. Accessed on 1 December 2023 from <https://time.com/6246974/worlds-biggest-crises-protect-basic-needs/>.
- Zannettou, S., Caulėeld, T., Setzer, W., Sirivianos, M., Stringhini, G., Blackburn, J., et al. (2019). *Who let the trolls out?* in Proceedings of the 10th ACM Conference on Web Science - WebSci '19 (New York, NY: ACM). doi: 10.1145/3292522. 3326016
- Zeitsoff, T. (2017). *How Social Media is Changing Conflict*. *The Journal of Conflict Resolution*, 61(9), 1970–1991. <http://www.jstor.org/stable/26363973>
- ተክላይ ገብሩ. (November 10, 2022). *ምን እያለ ነው ይህ ሰው* (Video file). Accessed from <https://www.facebook.com/1357285723/videos/1487984155005519/>

Appendices

Appendix 1- Interview Guide

አዲስ አበባ ዩኒቨርሲቲ፡ ትምህርትና ስነ-ባህሪ ኮሌጅ

የሳይኮሎጂ ትምህርት ቤት

የፍቃድ መጠየቂያ ሂደት/ፎርም

የተሳታፊ መረጃ

ኮድ	እድሜ	ጾታ	የት/ት ደረጃ	የስራ መደብ	በስራ መደቡ ያሉት ልምድ
1					

መግቢያ

ስሜ ታደሰ መገርሳ እባላለሁ። በአዲስ አበባ ዩኒቨርሲቲ ሳይኮሎጂ ት/ት ቤት የፒኤችዲ እጩ ስሆን በአሁኑ ሰአት የመመረቂያ ፅሁፌን በአንላይን የጥላቻ ንግግር ላይ እየሰራሁ ነው። በዚህ ጥናት ውስጥ በቁልፍ መረጃ ሰጪነት የመረጥኩህ ለጉዳዩ ቅርብ እንዳለህ በማሰብ ነው። በዚህ ቃለ መጠይቅ የምትሰጠኝ ሀሳብ ለጥናቱ መሳካት የበኩሉን አስተዋፅዖ ያደርጋል ብዬ አምናለው። የእርስዎ ስም ወይም ማንኛውም መለያ መታወቂያ ስለማያስፈልግና ከሰጡኝ መረጃ ጋር ስለማይያያዝ በተሳትፎት ምክንያት ምንም አይነት ጉዳት እንደማይደርስብ ለመግለጽ እፈልጋለሁ። ነገር ግን በሪፖርቱ ውስጥ የርሶ ሃሳብ በአግባቡ የሚገለጸው ቃል በቃል በማስቀመጥ ብቻ ሆኖ ከተሰማኝ ስምትን ሳልጠቅስ ሀሳቦችን በጥቅስ ደረጃ ላስቀምጥ እችላለሁ። ይህ ጥናት ሲጠናቀቅ እርስዎ ከጠየቁ የጥናቱ ውጤት እንዲደርሱት አደርጋለሁ። ይህንን ጥናት በተመለከተ ተጨማሪ ማብራሪያ ከፈለጉም በኢሜይል noh16509gmail.com ወይም በስልክ 251911060753 አጥኚውን ማግኘት ይችላሉ።

የሚከተሉት መሪ ጥያቄዎች ሲሆኑ ለመመለስ አስቸጋሪ ወይም ተደደጋጋሚ ጥያቄ መስሎ ከታየት ጥያቄውን መዘለል ይችላሉ። ከተቻለም ለጥያቄዎቹ ሰፊ ያለ ሃሳብ እንዲሰጡኝ በትህትና እጠይቃለሁ!

1. ብሄር ተኮር የጥላቻ ንግግሮች የትኩረት ቦታዎችና ምንጮቹ ምንድን ናቸው ብለው የስባሉ?

2. በርሶ እይታ ሰዎች ለምን ጥላቻ ንግግር ውስጥ ይገባሉ/ይሳባሉ?

3. የጥላቻ ንግግር ስርጭትን በተመለከተ የማህበራዊ ሚዲያ ተጠቃሚዎችን ባህሪ እንዴት ያዩታል?

4. በኢትዮጵያ ውስጥ ብሄር ተኮር የጥላቻ ንግግር እንዲፈጠርና እንዲቀጣጠል ከማድረግ አንጻር መዛግባት/አከራካሪ የታሪክ ድርሳናት (texts/books) ማግኘት ይቻላል (x§cW BIW ÆI#) እንዴት?

5. ብሄር ተኮር የጥላቻ ንግግር እንዲፈጠርና እንዲቀጣጠል የመሪዎች፣ የምሁራን፣ እና የዩኒቨርሲቲዎችን ሚናን እንዴት ያዩታል?

6. በርሶ ምልክታ ለጥላቻ ንግግር የተጋለጡ ብሄሮችን ከግምት ውስጥ በማስገባት (ካሉ) የጥላቻ ንግግር ስለባ ለምን እንደሆኑ የሚሰማዎትን ግምት ያስቀምጡ

7. የጥላቻ ንግግርን ለመቀነስ/ለመከላከል ያሉት አሰቻይ ሁኔታዎች ምን ምን ናቸው ብለው የሰባሉ?

8. በርሶ እይታ የጥላቻ ንግግርን ለመቀጠል የሚደረጉ እንቅስቃሴዎች ላይ እንቅፋት የሚሆኑና የጥላቻ ንግግር እንዲስፋፋ በር የሚከፍቱ ጉዳዮች የትኞቹ ናቸው?

9. የጥላቻ ንግግርን ለመቆጣጠር፣ ለመቀነስ ማን ምን እንዲያደርግ ይመክራሉ?

ስለጊዜዎትና ሀሳቦት አመሰግናለሁ!

Appendix 2- Sample Transcript of a Hatespeech Video

- File name:V15
- Video Time: 86 minutes
- Transcriber: Tadesse M .
- Key: P1= the first person, p2 = second person, P3= third person, [P2 and P3 talking about a P1's video]

NB: At some places, Ethnic names are shared with black color

P1- ████████ you have to be ashamed of yourself, it is disgusting. It is very difficult to live with you; you date your own children. Despite your shameful experience, you demonstrated on the street saying 'Ga*las couldn't rule over us'/we don't need to be ruled by Gallas. If you are honest and brave leave Oromia and go to your own Gondar and Gojjam; Oromos never cross your region. It you ████████ who never cease going and living in every land of Oromo. You are claiming Finifine city belongs to you; what are you doing there? You know who Qeeros and OLA are? ████████ activist are so shameless! ████████ you are criminal. All Kush descendants living in America should take ████████ to court as he is a genocider, who said 'it is only 70 SHANKILAS who killed'. The ████████ are claiming the responsibility that they killed. You mocked on the lives of black people and insulted them Shankilas! What a shame! Is this what you call being ████████; are ████████ dirty like this? Are ████████ mindless, not able to think? Who the hell is ████████? Is that those speaking on FB? Or the people living in Ethiopia? Or the Government there? You were full of louse, were hunger strike people in Ethiopia, but after coming to foreign land you unloaded your problems, changed your shorts and play trick on the poor Gumuz people. You are claiming Barara/Addis Ababa/..... You don't believe in dialogue and reality/evidence. You don't like to see leaders from other ethnic groups; you only want to grab power to yourself. You are liars [turinafa] who use almost all the media platforms in that country; you are beggars. If possible it is good to build walls against you not see you again. No social contact is needed with you, no exchanging goods, and living with you as neighbor. Ethiopia is cursed because of you! You are killing smaller children like the Gumuzs'. Why are we crying for Rwanda? The same sheet is committed by ████████ in Ethiopia. Don't you worry that other ethnic groups will stand firm against you? Have you ever thought of possible revenge from other ethnic groups? Where do you escape, Oromos, Tigress and Afara are all against you! You tried attacking the Kereyu people and received heavy punishment. You tried the same with

Wollo and Tigray people. You go here and there and attack/instigate others in their home. Your elites are sources of false history, drama /script writers, and propagandist. All Ethiopian are happy if you leave their territory/come out of Ethiopia/. You were shouting that your people were evicted from Burayu, then three years back you went on to the Wollo Oromo and attacked them. But they fought you bravely to extent that you again cried as a child. Then you shouted on Addis Ababa condominium because you were denied. You know what? You have shown what [REDACTED] is for this generation! That is you have 100% proved [REDACTED] cut breast, hand, that [REDACTED] are genociders, that they kill smaller children, that [REDACTED] are those who exterminate others identity such as others language! You killed men and say ‘we killed SHANKILAS’! those who give birth to [REDACTED] is bad.

Who said Amharic is Amharas’ only? It is the language came from Armania. Orthodox Christianity and the language came from the Armania! You are settlers/METE/. The land of Gojjam and Gondar belongs to Agew and Qimant respectively; not yours! You may not know your fathers as you may be DIQALA! You become AZMARI, beggars everywhere, killing the minority who have no power to defend themselves. I used to believe the rural communities of [REDACTED] are naïve/innocent; but now I have seen how cruel they too are like their masters in towns/cities! The You merely seek the lands of others, shouting for Diredawa, Finfine, Wollo, and now Gumuz’s land! You shout for it but in vain! [REDACTED] is a criminal who posted the dead body of smaller children of Gumuz killed by Amharas! It is not in the social media that bravery counts, [REDACTED] you should go join ABIN and fight along with them there. [REDACTED] are coward, psycho!

As Gumus children are still being killed, we must be voice to them! You instigated Oromos, Afar, Tigray and they punished you. Then you found a soft /weak people/Gumuz and attacked them. You are coward enough! They are innocent, poor, weak, and unorganized. We Oromos were smaller than you are the time you committed Genocide on our ancestors! You know call Gumuz Shankila in the same manner you were calling the South people ‘slaves’. Remember Ethiopia exists because of Oromos resource and development! If you leave all us, the rest of ethnic groups will live in peace with noble Federalism!! You complain that [REDACTED] are displaced from Oromia. Who brought you to Oromia in the first place? It is because of famine? You came to Oromia land because of starvation, and for your survival! Remember, to Amharas, Oromia is like Paris or London or DC where you can see good people, things and ways of life!! Your life style, culture and thinking are so disgusting. There

Insult,
devaluation
Mockery to
subgroups
misfortune
ethnoreligio
hate

is no ethnic group in Ethiopia that escaped the abusive and genocidal acts under the Green, Yellow, and Red flag.

P2= today I shared this video not for political reason but to let those Amhara politicians who are divided along Gondar, Gojam, Wollo etc kow what our enemies are thinking for us! This woman has told us the politics they arranged for Amharas clearly! For example, they are claiming Wollo to take it to Oromia (Kemise Oromo music opened here as prove for the claim). There are Amhara politicians who never want to hear about a united 'Amhara'; and that those Amharas outside the region are condemned for saying 'one Amhara'. But we (Amharas in the diaspora or outside the region) are saying the need to have a country for Amharas (Named father land Amhara) because we need 'our homes' [METELEYA/TAZA] in time of trouble. [REDACTED] (the previous speaker), is reading from notes if you closely looked her. That means she was given agenda to air! Some of you say Misrak is naïve woman as Jiji Kiya is. But remember it is these individuals who were ushered by police while they were in Ethiopia. They were given agenda there at Ghon Hotel. Let me give chance to Yordi, you can continue

P3= I feel pity that we never learn or take lesson from things or our enemies. They have around 16 parties but one goal! If they see unpopular idea from you, they will attack you together. Even Amhara artists like Amsal mitike is ignoring the Amharas unity.

P2= ok guys, you have heard what our enemies are doing! She was give agenda by her masters. Many times when they speak/discuss in their language is the same as what Misrak is telling us now. What Dr. Abiy is telling us about Ethiopia, MEDEMERA, and 'unite' is fake and behind the curtain this is the main mission they all have! they [REDACTED] are wishing even the land of Gondar! They have already prepared a map in which they included parts of Gojjam and Wollo to their territory! They are trying to torch us psychologically by denying our identity and language. Hence, the Amhara youth should avoid divisive lines and discourse and unit yourself to mitigate what your enemies are digging/preparing for you!

-----End of sample transcript -----

accusatio
in a mirro

Appendix-3 Sample Comments analyzed

ቆሻሻ ድንጋይ

ግን በታሪክ የመጣችሁበትን የት እንደሆነ ታውቃላችሁን? ታሪክ የላችሁም ያወረብ ዘር ሰለሆናችሁ ሀገሪቱን ለመቀራመት ለመሸጥ ካልሆነ በስተቀር

በጦርነት ጊዜ ጀለብያችሁን እየከሳችሁ ወደመጣችሁበት

Nebese segashi yegenetatel

የሰው ልጅ እንደዚህ ሴቶች ደደብ ሆኖ ይፈጠራል።

ከፍት አፍ

ጋ*ላ

Kmalamu man edehone tawukiwaleshi botiyam

እር ይቅርበልሽ እር እ/ር ምን ፍጥርት ነሽ ጂቦን ይጣልልሽ አችን አይሉ ላጉች ስትናገሪ ውድ እምትመስይ ጋ*ላ ነሽ ምን ልታመጪ ነው

መጀመሪያ ዝብሽን ባባረሽ ኢ.ኢ. ላሃጫም ቱ ምን ትመስያለሽ በእግዚአብሔር ቱ ገልቱ

ምን ትመስላለች ጎተራ ፊት ስደብሪ ዘግ ሀይዋና አጫሉን የበላልሽ የረሀስሽ ሰው ነው ጋሎ

ከብት እነዚህ ሰዎች ግን መች ነው እደሰው የሚያስቡት

ኢትዮጵያ ስትፈርስና ስትበተን ትግራይን ብቻ ለይተህ በስጦታ እቃ ጠቅልለህ ልትወስዳት ነው አንተ መሀይም

አንተ ራሥህ ገደል ግባ በወያኔና በመሠሎችህ የሞቱት የድመት ነፍሥ የገበሩ መሠለህ ገና በቁምህ ትነዳለህቆሻሻ ናቹ ገና ትፀዳላቹ ለኢትዮጵያ የማይጠቅም ሁሉ አጋሥሥ

እግዚአብሔር ይገስጥህ ኢትዮጵያን እግዚአብሔር ይጠብቃታል አንተ ኢሉሚናት

እግዚአብሔር፡እንኳን፡ከተዛተ፡ከተወረወረ፡፡ያድናል፡ቢሆንም፡ያበደ፡የሚሰራውን፡አያውቅምና፡መጠንቀቅ፡ጥሩነው።

Libona yseteh

አይንህ ይጥፋ አተ ደደብ

ምንም የበይ ተመልካች መሆን ነው

ሻሻመኔ ዘዋይ አጣዬ ያቃጠለ መከላከያ ልዩ ሃይል ነው አማራን በመርዳት ሥም ዘበት ነው

ትክክል

አማራ ምን የሚባል ዲሲፒሊን ያለው ህዝብ ነው

ወናፍ በለው

ሸረ ብዙ ነው ጉዳ እኛ ላይ ያልተሠራ ህይወት የለም

አካራ መዋጋት አይችልም አህያ ደንቆሮ ነው

ቤተልሄም ሻዕቢያ ነች።

ዳዊት አርአያ ሕወሓት ሥትወርድ የሥም ጭብል አብሮ ወረደ□□

ኢትዮጵያ ውስጥ በብዛት አርቲስት prostitutes የሐረር የግላ ወ ጋላክሲ ዲ ዎች ናቸው

ሸረ ፎጣ ለባሽ ህል ተደምጧል

ትግራይ ፀረ ተዋሕዶ manifest የቀረፀ ፅዎናዊ□□

ትህነግ እየዘረ ጡት ቂ* እያለ እያባላ ይበላ ነበር።

እና 1983 ዓ/ም ጀምሮ ማን ነው ነፃ ያወጣቸው? አንተ ራስህ መፃፍ ማንበብ ያቻልከው እኛ በከፊልነው መስዋእትነት ነው።ሾርት ሚሞርያም የናንተ ተረት ተረት ነው

የሰሜኑ ምሽግ ይሰበር እንጂ ሌላው ዕዳው ጉበስ ነው የምትሉት መሬታቸው ዘርፋቹ ቋንቋቸውና ማንነታቸው ለማጥፋት አይደለም እንዴ?

አማራን ከወያኔ ባርነት ነፃ ያወጣው አሮሞ ነው። የአሮሞ ህዝብ በቆራጥነት ወያኔን ታግሎ ባያስወገድ ኖሮ አማራ የወያኔ ባርያ ሆኖ ለዘለቄታ ይኖር ነበር ! በመርህም ዋጠው !!

ደገ አረጉ እንኩአንስ አሉ

የውሻ ልጆች በሀገሪቷ ሀብት ጠግበው ምን ያርጉ ታዲያ

እባክህ ሙልዩ እንዲህ አይነት ነገር በዚህ ወቅት ተው ። ማለቴ የማውቅልህን ተሥፋና ሞራል ለግስ

እረ ኦሪጅ ይታረድ

ትግራይና አማራ መታረቅ አለባቸው።

ፀሎት ብቻ

አንተህ ዳቢሎስ ነህ እውነተኛ ዲያቆንማ የእ/ር ቃልን ነበር የሚሰብከው ዳሩ የበግለምድ የለበስከ ሰይጣን ነህ ማፈሪያ

ደረጃ ወልዱ ስንት ላፒስ ነው የዋጥከው??

ባሠቃቂ ግድያ ንጹህን የሚያልቁባት ሀገር ባለም ባሁን ሠአት ኢትዮጵያ ብቻ ናት ዳዊት ታላላቅ ሠዎችን አሠባሥበህ ይህንን የተደፋፈነ ነገር ላለም ግልጥ አርገው ድከም ልፋ በመጨረሻ እድሜህ ይህንን ታላቅ ስራ ሠርተህ እለፍ አደራ አደራ አደራ

Aamara negn mesrtie Ethioipiawe

ባንዳ

እውነት ነው

ኦብይ አምሮ ዩለውም ልክ ነው

ኦብይ መውደቁ አይቀርም

Look enezih yeminilik yeshintamu yekomataw lejoch kkkkkkkkkk

ዘመን ሲበላሽ አዝማሪም ጋዜጠኛ ይሆናል

ዳዊት እድሜውን በመበጠበጥ የፈጀ ነውረኛ ነው

እናንተ ጅቦች ኦይ ኦይ አትበሉ።

የደሀ ልጆች አስጨረሳችሁ ነብሳችሁ አይማር እናተ ልጆቻችሁ ውጭ ናቸው አናተ ጥሩ መብላት መጠጣት ይችላላችሁ ፋጣሪ አዲ ይፋረዱ ።

ፋጣሪ ይች ምዲር ቀሪ ናት ግን አተ በጣም ትገርማለህ እውነት አተ ፋጣሪህን ምን ብለህ ትለምን ይሆን ከፋነት አይጠቅምም ዋ መጨረሻህ

ኢትዮጵያን የማይወድ ባንዳ ከጣሊያን እስከ ግብፅና ግብራክበሮች ኢትዮጵያን ሲሸጥና ሲበታትን ቆይቶ ኢትዮጵያዊ ነኝ ማለት ስላቅ ነው።

መጀመሪያ ሳየው ደግጬ ነበር ፊደሎቹን በደብ ሳይ ፊክ መሆኑን ተረዳውአንተ ብቻ በርታ አጫጫውማ ይህን መጋ

እኔ ያንተ ፔጅ አለመሆኑን መርምሬ አግንቸዋለሁ ግን ብዙ ተከታይ እንዳለሁ ተመለከትኩ

ሁሉም ፔጆች አንድ ሰው የፃፉቸው ነው ማመስለው

አንታለልም መምህር የአጳጵፍሩህን ዘይቤ እናውቃለን

ለሱ መልስ መስጠት መውረድ ነው

ንገርልኝ

ግን በጣም ብዙ ግዜ ተስዳድብን ምንም አለመጣንም

ፊፋኤል እራሱን ተቆጣጠር

መንግስትነው፣ የዝሀሁሉ መልስኤ፣ እግዝያብሄር፣ ይፋርድበትበስ

Appendix 4- Sample Pictorial Posts



Yekoye Lemma and 12 others · 4 comments

Like Comment Share

- 

Shito Tulu
666toche nachawe

Like Reply
- 

Zewdu Gebeyehu
እይይይ እነሱ ከቆሻሻ ያንሳሉ ቆሻሻ ቢያንስ ማዳበርያ ይሆናል

Like Reply

Voice Tigray · Follow
1h · 🌐

#ንታሪክ_ይቀመጥ!!

ድለይት ፀላእቲ ኣብ ልዕሊ ህዝቢ ትግራይ እዚ... See more

ጋዜጠኛ ማክሴል ነኝ Emirates D... · 13h
#ይገቡብ!

"እጅግ መሰሪና እውዳሚ #ትግሬ የሚባል ከፉ #ህዝብ ነበረ። ከአማራ ከአፋርና #ከአርትራ ጋር ባይረገት ጦርነት ከምድር ተጠራርገው ጠፉ። ከ2016 በፊላም ከቶውኑ አልታዩም" የሚባልበት ዘመን ይመጣል!

© ኑብዩ አሙሃ

Urbanization or Supremacy? Unpacking the Oromummaa... See more



borkena.com
Urbanization or Supremacy? Unpacking the Oromummaa

Oromia Chronicles · Follow
January 6, 2023 · 🌐

ከነት ያገኘው አብዮት (literally) አይ ተዋድርስ በእርሞ እና እስልምና ጥላቻ ተነሳሰቱ የፈፀመው አረመኔያዊ እልቂት (በደብዳቤው ላይ 'አንድም ላልተው ፈጀዋቸው' ይላል) ኢትዮጵያን አሳኝ እስላሞቹ ከላሉት የውበት ትርክት ጋር ስለሚጋዩ ... See more



ክትባት የሚያስፈልገው ዘር ተኮር ጭናጨፋ ለሚፈጽመው የአብይ አህመድ/አህዲድ የአሮሙማ አገዛዝ እንጂ ለአማራው ህዝብ አይደለም!... See more

TST APP 6h · 0

Amhara Media Corporation/ አማራ ሚዲያ ኮርፖሬሽን 45m · 0

ከ3 ነጥብ 5 ሚሊዮን በላይ ዜጎች የኮሮና ቫይረስ መከላከያ ክትባት ለመስጠት ማቀዱን የአማራ ክልል ጤና ቢሮ አስታወቀ::... See more



ጌታ ሆይ እትመጣም? ወይስ ደረጃ ይምጣ? ጌታ ሆይ እንደውም ደረጃ? ወይ ቀልጥ መልስንል እና አክሱም? ወላጋን ያስተዳድርልን:: ከደፊ ውጪ እየሰ ፀ... See more

← Post

DerejeK68472577 @DerejeK68472577

አማራ ሆይ በነካ እጅህ አክሱምን ወደአማራ ክልል የመጠቅለያህ ጊዜው አሁን ነው:: ወላጋን ወደአማራ ክልል የመጠቅለያህ ጊዜው አሁን ነው::

[Translate post](#)

Rashid Abdi @RAbdiAnalyst - 2h Fano militia gunman takes selfie against background of slain Oromo.

Note Ethiopia flag jacket and Christian cross.

Fano extremists see themselves as enforcers of God's justice, protectors of "One Ethiopia". They are the perpetrators of the worst violence in Ethiopia today.



Appendix 5-Unstructured Sample Memos

- it seems that a person sharing hate speech doesn't want any response that argues against that message. may other hate mongers on SM exist than this type?
- it is known that deferent animal names, weeds, and the like are assigned to groups. what is the connotation? why Ebab, Misit, Kebt,? does the connotation of this animals portrained in other narratives or accusations and insults expressed blatantly?
- I have decided to include any provocative and hate fueled claims as hate speech. such thing include an attempt to explain historical accounts in a way that others perceive as hate speech. this of course is ruled out in the definition of hate speech adopted by Ethiopian government-'any scholarly discussions cannot be taken as hate speech.' I have decided to take them as hate speech because of the recipients' perception and reaction.
- A hate speech or hate fueled post is seen to attract more discussion like and various types of comments if the post come from a high-profile name' person or a scholar; while those hate speech post whose source are perceived to be lay persons will beget a homogenous response mainly attacking the source of the speech. More like-minded following that person; layperson followed by their equivalent? laypersons have no idea to articulate and defend? try to find possible explanations for this insight.
- if a hate speech or fake news gets viral and is found to be hacked accounts, should the data be included in the analysis?
- the fact that videos produced by unknown/lay persons but saved by public figures are used in this study because of reasons. why they saved it? because it has impacts
- can I add a research question that seeks to identify scholars assisted discussion/agendas that attracts hate speech. that would help me specifically focus on what is shared by scholars in Ethiopia. for example, how the many but broken hate speech replies and comments are found in the scholars debate and narrations?
- Is there 'Nvivo hate speech' which takes the exact name of a particular ethnic group as insulting or derogating word. an instance of this type is ደደብ ጋይንቱ ነገር ነሽ::
- ✓ shall quotes be translated or put in the exact language they are found?
- ✓ how to make sure the truthfulness of the fb account- whether they belong to the name they bear.
- ✓ የቄሮ መንጋ በፋኖ መንጋ ላይ መንገድ ዘጋበት ሁሉም መንጋ አመድ በዱቄት አሉ; ባለ ነጠላ ጎጃም ሁላ; የቆምጨዩ, ባለግዜዎች, is this hate speech, How?
- ✓ interpretation was possible during data collection, screening, and storing.

- ✓ starting with hate speech and ending with praying as indicated in the following extract:
- ✓ I feel some are hard to be judged as hate speech; yet they are more likely to attract hate fueled discourses and animosity.
- ✓ implied strategies of the sources of hate speech are less likely to be inferred from the comments and replies, but from the main post, especially the ‘bigger’ narrations . Hence it is better to have a data set for the main posts
- ✓ shorter videos of hate speech are not preferred primarily because they could be taken out of the real context in which the speech is made.
- ✓ enjoying and celebrating death of outgroup elite. Why?
- ❖ some of the hate mongers appear to ‘advise’ the out-group, while showing them their weak sides and the need to correct them as soon as possible so as to avoid self-imposed negative outcome.
- ❖ When an individual insults an ethnic group, either he or his ethnic group, or both could be insulted back.
- ❖ Derogating regions without mentioning ethnic groups. such as ‘oromia is a region where የሰው ልጅ በደቦ አንድ አባብ በድንጋይ የሚቀጠቀጥባት
- ❖ fake news as a hate speech if it is aimed at derogating out-groups or make them vulnerable to attacks or abuse of some kind.
- ❖ I came to see the overall scenario analysis should be part of the context in which hate speech is framed and disseminated.
- ❖ There is also Intra-ethnic hate speech- Gondar-gojjam- wollo etc; Arsi, shewa, Bale, Wellega etc
- ❖ individuals action seems more responsible in triggering hate speech: a person says or does something – triggers his ethnicity- hate speech
- One magnetic area for hate speech is narrations which have no empirical evidence, are speculations, conspiracy reading.
- how many people have seen that hate message? the number of likes, shares, retweets, replies etc. are indicators of the number of individuals who have ‘seen’ the message. yet, the number showing these indicators should be judged along with the date the message was posted.
- Sometimes, the hate speech comes not from individuals but from groups such as during demonstration or crowd/gathering/rally!

- It is found that the pejorative words used to insult others such as JUNTA, Neftegna, FOTE etc. are taken by the target of the speech and they used in as a sign of bravery or good symbol of their identity! a typical example for this could be ...I am a brave junta!
- Competing accusatory narrations: Maykadra massacre- accusing each other for it
- the alleged internal displacement and alleged killings in some places such as wellega, North Shewa[Ataye] and Metekel zone of Benishangul Gumuz remain a hot agenda for hate fueled discourse.
- Dual hate mongers: insult ethnicity, religion, individuals, etc at the same time
- dual targets: Oromos are targeted for being Muslims and oromos, and also supporting a party such as OLA. an attempt to divide the people ‘የሸዋ ኦሮሞ እራስህን ከአሩሲ አራጅ ኦሮሞ ለይ በስመ ኦሮሞ አትጨፍለቅ ጃዋር አይስስ ነው አትከተል እምነት የለውም እምነትህን ሊያስጥልህ ነው ጃዋር አዲስ አበባ የሸዋ እንጂ የአርሲ አደለም ፕሬድ ደጅህን ከወሀብያ ጠበቅ:
- Sometimes when a member does something negative offline, their member support them in the online platform, such as motivating their wrong doer. So, pacemakers could be dyad in online platform only and between offline and online aspect.
- Please bear in mind that all ethnic groups are nearly equally insulted by all the four types of hate speeches. Yet, there are appears to be skewed derogations in some of the expressions such as the use of animals.
- Hate speeches has cognitive, affective, and behavioral components.
- Can a psychological definition of hate speech be same as a legal definition? How? The purpose of hate speech definition from these two discipline seems different.
- Can I divide the dissertation into three studies, which unique methodologies for each? NO
- Some contexts of hate speech are immediate while others are distal
- The mechanisms for hate speech could also be seen as a means for the propagation of negative stereotypes and related negative intergroup attitude we observe on social media.
- The use of the word Ga*la is reported to be traced in the work of Aba-Bahren’s book- ‘Zenahu-le Galla’.
- This dissertation has to employ multi-level analysis: the main hate speech post, the comment, the context, and the dialogue going between users, as situation allows.
- Why ethnicity is mainly implicated in the hate speech? If Feyisa, and Hagos did something wrong, it is their ethnicity that is inferred than other social identities such as their religion.

- How can I make sure the pictures and videos used in this paper are deepfake [edited]. What is and what if not?
- I feel I have implicitly used ‘scenario analyses as a method to identify codes and possible concepts. This means, more than what is said, how it is said and the overall situation that include many users at a time.
- some users have the tendency to randomly react- like, share, or comment on every post they across with.
- Do I need to use a checklist to assess ‘likes, shares replies, views etc. of hate speech posts? I think this is necessary.
- Content analysis can not help extract roles from sharing or liking –this can be addressed via GT. It also never give insight about link among themes
- All poachers are trolls but not all trolls poachers.
- Because of the involvement of educated people in hate speech and related rhetoric, education and intellectualism is mocked at. Some say ‘ if the purpose of education is to be like this, we don’t need it.’
- Ingroups prefer to make a Hate speech posts targeting their ethnic group to go viral so that many would get it; file legal cases, or document it for history.
- Both outgroups and ingroups share hate speech for different purpose
- Objectification can also been inferred from actions. For example, burning dead body of ‘enemies’ imply the human being is treated as objects. This also shows, the group which burn dead body will be dehumanized
- It has to be noted that I omitted presenting some of the extraordinary offensive expressions, as I think they are more than enough in their loaded emotions.
- The fact that hate speeches can be made publicly such as in a meeting full of people and the hater can be applauded shows, the willingness to derogate others as well as anonymity is not a big issue.
- The analysis made takes different approaches. First specific contents are analyzed. This contents can be text or pictures. Other codes are emerged as I analyzed the overall situation, such as the conversation going on between two or more opposing individuals [discourse]. Still, codes can also be emerged as insights or reflections made outside the ‘field’ and then looked for evidence.
- Forgettable and Unforgettable hate speech! Unforgettable ones are either vivid, extraordinary, or made by elites Or they are sources from an extraordinary brutality!

- Contexts can also be hidden and surface level contexts.
- Hate speech is disseminated because of calls made by members to make it public, by choice of users, etc.
- Chronic Hatred for poachers
- Earned reputation is highly supported by Sternberg triangular theory of hate. Part of the description is ‘Disgust can be triggered by the characteristics or actions of the hated target’.
- Trolls/poachers can be thrillers, offending or defending/retaliatory in nature.
- Some of the hate speech expressions infer stable characteristics of the target group, as seen in the following quotes: የጡት ቆራጭ ልጅ-በልት እና ጡት መቁረጥ እማ ከዘርህ ነው

There are expressions which can not be judged as hate speech, but more likely to elicit hate speech as seen in the following excerpts: ወለጋ አማራ ነው; የተምቤን አገዎች ነፃ የሚሆኑበት ዘመን ሩቅ አይደለም; እናንተ ወልቃይት ትላላቹ እኛ ደግሞ አብርሀ ጅራ አብደራፊ እና ለማሊሞ እንዲመለሱልን ነው የምንፈልገው; ልሙጡ ባንዲራ ይወክለናል ሚሉት ኦርቶዶክስ አማሮች ናቸው

- A person would comment; God forgive them, they are just animals’ which category to assign this? this is more of role confusion.....
- While studying hate speech contents, it is necessary sometime to look how positive contexts or messages can bring hate speech.
- Trolls are identified from comments. As such their consistency and techniques are less likely to be inferred. So, the last research question helps to identify poachers with the help of the previous research questions as well as reflexive journals.
- Poachers technique: they employ techniques in which they just speak positive things about themselves, but to let others have no that attribute. Hatred and offending others via self-pride. At times, they employ a Barnum statement is a personality description about a particular individual or group that is true of practically all human beings.

The fact that Ethnic Oromos are vulnerable to hate speech because of PP is also seen in rumors like the ‘Oromuma Project’ or ‘Kegna Project’

- Hatred is traumatizing physically, emotionally and morally.
- In irritating target--- users may ignore the whole group and focus on individual
- **Illusion of Differentiation:** failure to understand an insult offend others beyond the actual target of the speech. users fail to assume the derogating work they use to irritate a particular individual also affects the entire group the individual is a part.

- The country never ceases agenda that can be used as context for hate speech. hate mongers always have pretext and context.
- A particular content taken out of context may not allow one to understand the content is hate speech. This means, contents have real meaning when understood in their context. this shows, while some are blatant hate speech, others are disguised and only be understood as such when taken in their context. ትላንት እንደ ዜጋ እንኳ ከማይቆጠሩ ሱማሌዎች አብራክ የተገኘው መስጠፋ ማንቂያ ደውላችሁ ይሁን ሞት ለመንደር ፖለቲካ apparently seems positive but has negative meaning as well [**confuser**]
- A content could be a single word, a combination of two, a phrase, a full sentence, a paragraph, or more than one paragraph. Lack of diversity tolerance is mainly observed in attacking Irreecha celebration, Geda
- It would be better to search for textual hate contents such as hate filled books written which describe a group as a historical victim while another as a perpetrator. Such document analysis could shed line specific contents as well as mechanisms used to enemize others.
- It seems that contexts determine the types of hate speeches to be framed. For example, in the context of the conflict in the north and failure to uphold rule of law, dehumanizing expressions and enemification is more prevalence, while in the context of Irreecha celebration, devaluation is more apparent. The lack of law enforcement begets more of the dehumanization
- The level of mistrust observed in the hate speech is extended to thwart help from members of the hated/suspected group. for example, የጋላን ድጋፍ አንሻም በአማራ ስም መነገዳችሁን አቁም የአነግ አሽከር
- the weaponization of public discourse for political gain is a common issue in the hate speech observed.
- Some users explicitly accept they have hatred against their target, as seen in the following quotes’ As proud Oromo I stand with Tigray ppl. In fact the hatred I have for Tigray ppl for what they have committed to my ppl surprised the hatred that I have for ethiopiast including sell out Oromo’s and so call Amhara in this past two years.’
- Some elites are also given nikenames such as ‘Mismare’ and ‘the Tetracycle’ etc because of the role they play. This also takes us to the idea of earned reputation.
- Some hate speeches are against ‘Ethiopia’ :-ኢ/ያ ሁላችም የሰጣን ዘር ናቹ; The priest who called Ethiopia Elzabel; Let us destroy Ethiopia

- Hate-Bridges: are individuals, subgroups, and ethnic markers that an ethnic group is mainly insulted for. The following table presents some hate-bridges. In this analysis, hate bridges are not inherently negative but hate mongers can use them as a gate way to offend the entire group. they have dual meaning-liked and disliked; may be legally acceptable and politically needed; groups' liabilities, weak sides

Valence: this subgroups are highly valued by some ingroups while some relevant outgroups strongly reject and hate them; Criminalized: even not all, some



This show most individuals view but not react to hate speech posts-bystanders

Appendix 6- **Hate Speech Screening Checklists**

S/N	Does the content (within its context)	Decision	
		Yes	No
1.	Potentially offend the target?		
2.	Describes target outgroups using pejorative, derogatory, or devaluing terms		
3.	Portrays targets in a way that exposes them to prejudice, discrimination or violence?		
4.	Does the content call for some kinds of action against target groups?		
5.	Describe targets as enemies, threats, or wish them suffering		
6.	Contributes to intergroup animosity, mistrust, conflict or hatred		