The Mechanisms and Consequences of Sharing Affective Content on Social Media



Jonas Paul Schöne

Kellogg College, University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2023

Supervisor: Brian Parkinson, PhD

Co-supervisor: Scott Hale, PhD

ABSTRACT

The internet and especially social media has become one of our new natural habitats that we use daily to communicate with our peers and to get the latest news. Similar to their offline conversations, people online frequently use affective expressions to share their opinions or feelings about current events. However, unlike face-to-face co-present interactions, online content can reach a larger audience simply by being shared with followers with a click of the mouse. Since social media can reach such a wide audience, it is crucial to understand the types of content that users create and share and their potential beneficial or harmful impact. Despite the prevalence of social media in our daily lives, these questions have yet to be answered conclusively. Hence, the primary objective of my thesis is to address these gaps by examining the prevalence of affective content and its implications for expressing and viewing such content. Chapter 1 of this thesis provides an overview of definitions and measurement methods relating to online affective content, a review of previous research that aimed to answer the two questions raised above, and a discussion of the remaining research gaps that will be addressed throughout this thesis. In Chapter 2, I present an investigation into the kinds of affective content that are shared more when produced by public figures and ordinary users and assess which user characteristics influence the extent to which these kinds of affective content are shared. Building on these insights, Chapter 3 delves into the question of whether negativity is more prevalent than positivity in all political events and what specific language features contribute to the sharing of certain kinds of affective content. In Chapter 4, I aimed to address both of the overarching questions by investigating the types of affective content created and shared during an ongoing negative, but non-political event. Additionally, I explored the potential affective consequences of sharing such content on social media. To explore the impact of exposure to affective expressions on social media, Chapter 5 delved into the process of perceiving sequential

affective expressions as they appear on social media. Finally, in Chapter 6, I synthesize my findings with relevant literature and develop a comprehensive framework to explain the process of exposure to affective content on social media and its consequences. Furthermore, I discuss potential future research directions to advance our understanding in this area. Overall, this thesis provides insight into the types of affective content that are more likely to be created and shared in various contexts and by different content producers. In addition, it provides functional explanations for why people might share affective content on social media and explores the potentially harmful consequences of sequential exposure to such content on these platforms.

Für die Familie Schöne

ACKNOWLEDGEMENTS

Writing these following pages might have been one of the most challenging tasks in the completion of this thesis, that I did not dare to complete till the very end. The weight of responsibility to do justice to all those who have made it possible for me to sit here and complete my thesis, felt overwhelming.

To fully express my gratitude, it is necessary to revisit the expectations, or rather the lack thereof, that marked the onset of my academic journey. Growing up with dyslexia in rural Germany, merely keeping up with school was a significant hurdle, eventually culminating in my teachers expressing doubts about my ability to complete middle school. Given this dire outlook, the fact that I am now on the brink of completing a doctoral thesis feels nothing short of miraculous, a feat made possible by the individuals to whom these pages are dedicated. As such, this thesis is a story of *"US"*, motivated by the prospects of what people can achieve together with the goal to learn how to make collectives more united using the technology of our time.

Foremost, I extend my heartfelt gratitude to those who had to deal with me the longest, my beloved family. I knew I could have tried, failed, or succeeded in any career and you would have loved me the same way. Such unwavering support enabled me to embark on a path that led me to a role, a purpose, we had no knowledge of, yet perfectly suited me. Specifically, to my mom, you had to put in an unbelievable amount of work to get my brother and me to where we are, and I want to tell you that every success we have is a reflection of your own effort. To my dad, while you may forever be astonished over my lack of handyman skills and my tendency to work midnights, you encourage me to pursue a life that is so different from yours. Lastly, to my brother, though you may bear the playful title of 'prince' in our family, it's your leadership prowess, work ethic, and strength that truly crown you. I continually strive to emulate these traits and will forever look up to you, my older brother.

Before joining such a prestigious institution as the University of Oxford. I want to give due credit to another academic institution that was key to shaping my abilities. While Hochschule Harz may not enjoy the same recognition, it provided an incredible undergraduate experience I will forever cherish. In particular, I must express my gratitude to Prof. Georg Felser and Prof. Patrick Hehn. Their infectious enthusiasm for psychology and market research, as well as their effective teaching methodologies, easily rival those found at the top of the world. Given the chance to select my undergraduate institution again, I would proudly opt for Hochschule Harz.

In the same vein, I extend my profound gratitude to those who ushered me into the world of academic research, provided mentorship, and imparted the practical research experience crucial for embarking on a graduate degree. Prof. Stefan Sütterlin and Dr. Ric Lugo deserve special mention, who embraced me into the academic research sphere despite my lack of prior experience. I am indebted to Prof. Karl Steptoe for offering me the opportunity to work with Leicester City Football Club, illustrating how I could merge my passion for sports with an academic career. Prof. Mark Cropley, who welcomed me at the University of Surrey, was the first to nurture my ambitions to pursue a PhD. Last but certainly not least, I would like to thank Prof. James Gross, who has been instrumental in my development since our paths first crossed. His generosity in providing guidance and mentorship has been nothing short of remarkable.

Turning to the University of Oxford itself. The past four years have been the best years of my life, and were I given the chance, I would happily invest another four years into the PhD program. Oxford, unlike many institutions, fosters connections that extend beyond academic collaboration to personal relationships. I’ve never felt as integrated into a community as I have at this university.

Now, to those who made Oxford home. Maja, my academic sister, while our food preferences might span the distance we two ran together in the past four years, I rarely made a friend I felt closer to. Andreas, your talent for finding yourself in more drama than I ever faced always managed to put my stresses into perspective, but all jokes aside, you were carrying both Maja and me through the pandemic. Cody, our friendship was an unexpected delight; your class, style, and sophistication provided a welcome contrast to my simple tastes. To my formidable team—Ryan, Marco, Abdullah, Yanick—we not only played a sheer endless number of games together, but we also build a community now exceeding 350 members. Zaffer, despite your tenure at Oxford only spanning a year, people can hardly imagine one of us without the other. Marin, you endured the misfortune of being my housemate, yet you were still able to be my friend. Without your cooking and life advice, I am not sure how I would have managed my time here. And Yän, my best friend and partner in crime. Ever since we met at Stanford over five years ago, you have been a ceaseless source of inspiration, and I eagerly look forward to continuing our journey where it first began.

I would also like to thank those whose generous support made possible both my time in Oxford and the preceding education necessary for pursuing a graduate degree. I am immensely thankful to the Studienstiftung des Deutschen Volkes and the people of Germany, who have been funding my academic career since my second year of undergraduate studies. This scholarship has not only provided financial assistance but has also connected me with invaluable networks and offered me professional support, which has been instrumental in my academic achievements as a first-generation student. As an alumnus, I eagerly anticipate the opportunity to reciprocate this kindness by supporting other students who benefit from this scholarship in any way I can.

How can I thank you, Amit? When you were just finishing your own PhD, you have been my mentor, supporting me through every step, from the application process to securing

a post doc position. Although becoming an assistant professor must have been a tremendous amount of work and pressure, you found the time to tirelessly teach me, and I cannot express how much this meant to me. Seeing how we have built and grown your lab as well as carved my academic career out of, what must have been granite, fills me with joy. I am proud to be your first graduate student, and I hope these were just the first five years of many.

Last but certainly not least, to my primary supervisor, Prof. Brian Parkinson. You faced the challenging task of channeling my boundless enthusiasm and eagerness for exploration into careful thought and planning. Whenever I needed last-minute recommendation letters or feedback on my drafts, you graciously made time to support me. I deeply admire your exceptional ability to articulate complex ideas with clarity in your writing, and I am immensely grateful for your patience in helping me improve my own writing, which was no small feat. Most importantly, you provided me with the freedom to pursue projects that ignited my passion, while ensuring they contributed to my growth and progress. Although we may have appeared as an unlikely team, you successfully crafted me into an independent researcher with enthusiasm and depth.

SUBMITTED AND PUBLISHED MANUSCRIPTS

The research presented in Chapter 2 is presented in a manuscript presented by the *Proceedings of the National Academy of Sciences Nexus* journal and can be cited as follows: Schöne, J., Garcia, D., Parkinson, B., & Goldenberg, A. (in press). Negative expressions are shared more on Twitter for public figures than for ordinary users. *PNAS Nexus*. https://doi.org/10.31234/osf.io/wng5v

The research presented in Chapter 3 has been published by the Journal of *Affective Science* and can be cited as follows:
Schöne, J. P., Parkinson, B., & Goldenberg, A. (2021). Negativity spreads more than positivity on Twitter after both positive and negative political situations. *Affective Science*, 2, 379-390.

The research presented in Chapter 4 is also reported in a submission to the *British Journal of Psychology*, which combines my study with related research conducted by fellow DPhil student Gloria WS Ma, and described in her thesis. This thesis only reports the research contained in the submission that I led, but refers to the findings of the study conducted by Gloria Ma in its discussion. The submitted manuscript can be cited as follows: Ma, G. W. S*., Schöne, J.P.*, & Parkinson, B. (2022). Social sharing of emotion during the collective crisis of COVID-19. Available at: https://osf.io/9p3wh/
*Equal Contribution

The research presented in Chapter 5 will also be reported in a manuscript that is currently in preparation for submission to an academic journal.

TABLE OF CONTENTS

ABSTRACT ..... **I**

ACKNOWLEDGEMENTS ..... **IV**

SUBMITTED AND PUBLISHED MANUSCRIPTS ..... **VIII**

TABLE OF CONTENTS ..... **IX**

LIST OF FIGURES ..... **XIV**

LIST OF TABLES ..... **XVI**

LIST OF ABBREVIATIONS ..... **XVII**

CHAPTER 1. INTRODUCTION **1**

ABSTRACT 1
INTRODUCTION 2
AFFECTIVE CONTENT ON SOCIAL MEDIA 5
Assessment of Affective Content *6*
Engagement with Affective Content Online *8*
Prevalence of Affective Content Online *10*
CONSEQUENCES OF NEGATIVE EMOTION EXPRESSION ON SOCIAL MEDIA 16
Consequences of Sharing and Expressing Affective Content for Senders. *16*
Consequences of Consuming Affective Content for Perceivers *18*
OVERVIEW OF THE CURRENT THESIS 24

CHAPTER 2. NEGATIVE AFFECTIVE CONTENT IS SHARED MORE ON TWITTER FOR PUBLIC FIGURES THAN FOR ORDINARY USERS **27**

ABSTRACT 28
INTRODUCTION 29
Sharing Negative Affective Content on Social Media *30*
Public Figures and Ordinary Users *32*
Research Questions and Goals *33*
STUDY 1 AFFECTIVE CONTENT PRODUCED BY AND SHARED OF PUBLIC FIGURES COMPARED TO ORDINARY USERS 34

Method *34*

Results *40*

Discussion *53*

DISCUSSION 54
Limitations and Future Directions. *56*

CHAPTER 3. NEGATIVE AFFECTIVE CONTENT IS SHARED MORE THAN POSITIVE AFFECTIVE CONTENT ON TWITTER AFTER BOTH POSITIVE AND NEGATIVE POLITICAL SITUATIONS. 59

ABSTRACT 60
INTRODUCTION 61
Affective Content Sharing in Response to Political Situations on Social Media *62*

Research Questions and Goals *65*

STUDY 2: AFFECTIVE CONTENT SHARING FOR USERS WITH POSITIVE OR NEGATIVE EVALUATIONS OF THE SAME SITUATION 66
Method *66*

Results *70*

Discussion *79*

STUDY 3: AFFECTIVE CONTENT SHARING FOR USERS WITH POSITIVE OR NEGATIVE EVALUATIONS OF TWO SEPARATE SITUATIONS 79
Method *80*

Results *81*

Discussion *87*

GENERAL DISCUSSION 87
Implications of the Findings *88*

Limitations and Future Directions. *89*

CHAPTER 4. ONLINE SOCIAL SHARING OF AFFECTIVE CONTENT DURING THE COLLECTIVE CRISIS OF COVID-19 93

ABSTRACT 94
INTRODUCTION 95
Affective Social Sharing on Social Media *97*

Research Questions and Goals *102*

STUDY 4: AFFECTIVE SOCIAL SHARING DURING AN ONGOING CRISIS 104

Method *104*

Results *109*

Discussion *122*

CHAPTER 5. THE IMPACT OF NEWS FEEDS ON THE PERCEPTION OF AFFECTIVE NEWS ARTICLES AND OTHERS' RESPONSES 129

ABSTRACT 129
INTRODUCTION 130
Affective Content on Digital Media *131*

Aggregation of Stimuli Expressing Emotion *133*

Research Questions and Goals *136*

STUDY 5: EVALUATION OF EMOTION EXPRESSIONS IN TEXT SEQUENCES 137
Method *137*

Results *140*

Discussion *143*

STUDY 6: ANALYSING THE EFFECT OF THE AVERAGE EMOTION EXPRESSION ESTIMATION OF THE RESPONSES ON THE PERCEPTION OF THE CORRESPONDING NEWS ARTICLE 144
Method *145*

Results *147*

Discussion *149*

STUDIES 7 - 8: EVALUATING POSSIBLE EXPLANATIONS FOR THE OVERESTIMATION OF EMOTION EXPRESSIONS IN TEXT SEQUENCES 150
Method *151*

Results *153*

Discussion *158*

GENERAL DISCUSSION 159
Implications of the Findings. *160*

Limitations and Future Directions. *162*

CHAPTER 6. GENERAL DISCUSSION 167

GENERAL OVERVIEW OF THE CHAPTER 167
PREVALENCE OF POSITIVE AND NEGATIVE AFFECTIVE CONTENT ON SOCIAL MEDIA 169
Findings *169*

Implications of the Findings Relating to the First Main Research Question *173*

Limitations and Future Directions ... *176*

CONSEQUENCES OF AFFECTIVE CONTENT FOR SENDER AND PERCEIVER ... 180
Findings ... *180*

Implications of the Findings ... *182*

Limitations and Future Directions ... *186*

ASSESSMENT OF AFFECTIVE CONTENT AND ENGAGEMENT ... 189
Findings and Implications ... *189*

Limitations and Future Directions ... *195*

BROADER THEORETICAL CONSIDERATIONS AND FUTURE DIRECTIONS ... 200
Cultural Transmission Framework ... *200*

Future Directions ... *203*

CONCLUSION ... 206

REFERENCES ... 208

APPENDICES ... 249

CHAPTER 2: SUPPLEMENTARY MATERIALS ... 249
Analysis Using the Full Sample ... *249*

Sentiment Tool Comparison ... *256*

Topics Modelling Detailed Information ... *258*

Analysing Different Types of Verified Users ... *264*

Average Emotion Expressed by Users ... *265*

General Additive Model of Sentiment and Retweets for Public Figures and Ordinary Users ... *268*

CHAPTER 3: SUPPLEMENTARY MATERIALS ... 272
Emotional Language Intensity Predicting Likes in Studies 2 and 3 ... *272*

Emotional Language Intensity Predicting Retweets in Subsets for Positive Situations ... *278*

Emotional Language Intensity Predicting Retweets: Model Comparisons ... *281*

Emotional Language Intensity Predicting Retweets: General Additive Models ... *298*

Emotional Language Intensity Predicting Retweets: Sentiment Tool Comparison ... *305*

Analysis of most Viral Negative Tweets in Positive Contexts ... *311*

Removing Emotional Ambiguous Words from Same-sex Dataset ... *313*

CHAPTER 4: SUPPLEMENTARY MATERIALS ... 314
Linear Growth Models of Sentiment Change Over Time for Primary Social Sharing. ... *314*

Linear Growth Models of Sentiment Change Over Time for Secondary Social Sharing. ... *318*

Dictionaries *322*

Preregistered Pilot Study *325*

Baseline Period *326*

Repeating VADER Analysis Using COVID-Related Tweets Identified by the Long Dictionary *328*

Repeating All Analyses Concerning COVID-Related Tweets Identified by the Short Dictionary Using SentiStrength *333*

Preregistered LIWC Analyses *338*

Repeating the Granger-Causality Test Using 2 Weeks Instead of 1 Week as the Time Period *345*

CHAPTER 5: SUPPLEMENTARY MATERIALS 347

Study 9: Evaluation of Affect in Word Sequences *347*

Power Calculations to Estimate Experiment Sample Sizes *353*

Trial Structure, News Articles and Responses *357*

Amplification and Potential Moderators (Correlation with Scales) *364*

Race *365*

Notes. *365*

Gender *366*

Social Anxiety. *367*

Notes. *368*

Personality. *368*

Positive and Negative Emotional Reactivity. *373*

LIST OF FIGURES

Figure 2.1. Number of affective tweets by categories for both user types 41

Figure 2.2. The number of retweets as a function of sentiment and user type.....43

Figure 2.3. Differences between public figures and ordinary users. 47

Figure 2.4. Number of retweets as a function of sentiment and number of followers.....49

Figure 2.5 The number of retweets as a function of emotionality and proportion of political tweets. 50

Figure 2.6 Parallel mediation analyses the effect of user type on negative affect sharing via two different user characteristics. 53

Figure 3.1. Number of retweets as a function of emotional language used for events evaluated as positive and negative.....72

Figure 3.2 Number of retweets as a function of emotional language used for positive and negative events.....83

Figure 4.1 Sentiment Scores and Number of Tweets Posted During Baseline and Crisis Periods.....110

Figure 4.2 Frequency of Affective Content Types for Original and Retweeted Content 112

Figure 4.3. Number of retweets as a function of sentiment for original content during the crisis 113

Figure 4.4. Number of retweets as a function of sentiment for retweeted content during the crisis 115

Figure 4.5. Sentiment Scores of COVID-Related Tweets as a Function of Time and Users' Level of Pandemic-Specific Sharing.....118

Figure 4.6. Temporal Causality Between Sentiment Scores and Social Sharing of COVID-Related Content.....122

Figure 5.1 The comparison between estimated and actual average ratings of the sequence. 142

Figure 5.2 The comparison between estimated and actual average ratings of the sequence. 148

Figure 5.3 The likelihood to remember a response based on its emotional language intensity. 156

Figure 5.4 The likelihood to select a response as most representative based on its emotional language intensity 157

Figure 6.1 Cultural Transmission Model as Framework to Understand Affective Content on Social Media 202

LIST OF TABLES

Table 2.1. Breakdown of verified user types 39

Table 2.2 Associations between sentiment scores and number of retweets for different public figure types in comparison to ordinary users 44

Table 3.1. Most indicative retweeted unigrams and bigrams for both positive and negative affective tweet categories from conservatives in response to Donald Trump's election win. 78

Table 3.2. Most indicative retweeted unigrams and bigrams for both positive and negative affective tweet categories from liberals in response to the same-sex marriage ruling 86

LIST OF ABBREVIATIONS

Abbreviation	Definition
NLP	Natural Language Processing
LIWC	Linguistic Inquiry and Word Count
LDA	Latent Dirichlet allocation
EmoLex	NRC Word-Emotion Association Lexicon
API	Application Programming Interface
VADER	Valence Aware Dictionary and Sentiment Reasoner
PANAS	Positive and Negative Affect Schedule
CTM	Cultural Transmission Model

CHAPTER 1. INTRODUCTION

Abstract

In this chapter, I aim to shed light on the pivotal role that social media content consumption plays in its users' lives, while also addressing critical, yet unanswered, questions surrounding the type of content that is being consumed. Specifically, I argue that the use of affective language is a crucial driver of social media content consumption. However, due to methodological challenges in studying content consumption on social media, two broad types of questions remain unanswered: How prevalent is affective content on social media and what are the consequences for the sender and perceiver of such affective content? In this Chapter, I provide an overview of the various definitions for, and methods used to measure, affective content, and review the ways in which previous research tried to answer the two questions, as well as identifying remaining research gaps that need to be addressed within this thesis.

Introduction

During the German Revolution, the philosopher Ludwig Feuerbach coined the expression “A man is what he eats” (original quote in German: “Der Mensch ist, was er ißt”), which refers to the assumption that people’s consumption habits shape the way they think and feel (Feuerbach, 1862, p. 1). The scope of this phrase extends beyond the context of food to other types of consumption habits that might also influence who we are and how we feel, such as the consumption of news and other media. A seemingly ever-growing proportion of this consumption happens on social media platforms (Kümpel, Karnowski, & Keyling, 2015; A. Mitchell & Page, 2014). An average American adult spends roughly two hours daily on social media consuming different types of content and interacting with friends and strangers (Araujo, Wonneberger, Neijens, de Vreese, & Measures, 2017; Statista, 2022a). One major characteristic of these contents is that they are rich in affective expressions and valuation words (Bernabé-Moreno, Tejeda-Lorente, Porcel, Fujita, & Herrera-Viedma, 2018; Ferrara & Yang, 2015b; Steinert & Dennis, 2022). Such content I refer to as “affective content” in this thesis. Despite social media being an integral part of daily life and affective content being prevalent on these platforms, there are open questions regarding what type of content is produced and shared, why it is shared, and what the exact consequences are of exposing to and engaging with content that uses affective language.

Researchers have attempted to answer such questions, but in doing so, they have encountered both ethical and methodological challenges. For example, a controversial study on one of the largest online media platform, Facebook, claimed to provide experimental evidence that users’ exposure to content that uses affective language can influence their own emotional state (Kramer, Guillory, & Hancock, 2014). In this study, researchers manipulated the ratio of positive and negative messages that users saw on their newsfeed without their

knowledge. People who were exposed to more negative affective content subsequently expressed more negative affect in their own posts, and people who were exposed to more positive messages expressed more positive affect. This study not only provoked a storm of ethical controversies among researchers and Facebook users (Brady, Crockett, & Van Bavel, 2019; Goldenberg & Gross, 2020), but also highlighted how little is understood about the processes involved in the sharing of affective content online and about the consequences of expressed affect for users who are exposed to them. Although the researchers claimed that they measured these affective consequences, the experiment in fact only showed changes in users' communication without any further validation of affective state itself. This omission is a central limitation of purely digital media studies more generally.

Nearly two billion people worldwide are on social media (Statista, 2022b), which accounts for more than half of the English-speaking population of the world (Perrin, 2015). Given the large numbers of users, research and governments alike have raised the question of whether social media has a positive or harmful effect on individuals and society as a whole. Previous studies have reported mixed results, depending on how people use these platforms (for a review Kross et al., 2020; Lee, Lee, Choi, Kim, & Han, 2014; Luo & Hancock, 2020; Steinert & Dennis, 2022; Verduyn et al., 2015). When American users were asked this exact question, around 64% believed that social media has more negative than positive consequences for individuals and society (Auxier, 2020). One of the most cited reasons for this, in addition to cyberbullying and the spread of misinformation, is exposure to negative affective content. This observation seems to be particularly evident in political contexts (Brady et al., 2019; Rathje, Van Bavel, & van der Linden, 2021).

Users' belief that they are exposed to too much negative affective content can be influenced by two factors – the actual amount of negative affective content that they are

exposed to and their perception of the prevalence of such content on social media. To examine these two factors separately, we first need to quantify the negative affective content on social media by assessing how much negative affective content is created and shared on social media. We then also need to evaluate whether users are accurate in their judgment of the prevalence of negative affective content online, or if they overestimate how frequently such content appears.

The goals of this thesis are threefold. First, I aim to quantify how much affective content is created and shared on social media. Second, I seek to understand why certain affective content is shared in different contexts as well as how this sharing behaviour changes depending on who the creator of the content is. Finally, I examine what consequences affective expressions online have for users of social media and how affective content is perceived by users.

To help relate each chapter of this thesis to the current literature in the field, this introduction 1) outlines literature on different types of affective content, and how affect and content engagement is measured from social media text. I review the production and sharing of social media content, arguing that sharing depends both on the type of user who created the content and on the nature of content that is created (as elaborated on in Chapters 2 and 3). Then, 2) I review work on the perception of emotion on- and offline (as further discussed in Chapter 4). The final section of the introduction looks at the consequences of sharing and consuming affective content online.

Affective Content on Social Media

Research investigating the various types of online affective content uses a wide array of definitions that vary in the conceptual claims researchers make about such content. For example, the Facebook experiment mentioned above presented findings in terms of actual emotions, making implicit claims about the direct association between expression and felt emotions (Kramer et al., 2014), while the Natural Language Processing (NLP) literature uses the term *sentiment* (Hasan, Maliha, & Arifuzzaman, 2019). However, "sentiment" covers both attitudes and emotions, meaning that a reference to positive sentiment may mean that the user in question holds positive attitudes about the discussed subject (Hyvärinen & Beck, 2018; Medhat, Hassan, & Korashy, 2014) without necessarily experiencing any concomitant positive affective response to events. In addition to the differences in the claims regarding the underlying psychological construct of affective content, the terms "sentiment," "affect," "emotion," and "mood" are frequently used interchangeably to describe this content in the literature (Imran, Daudpota, Kastrati, & Batra, 2020; Kushawaha, De, Mohindru, & Gupta, 2020; Nahar, Sultana, Iqbal, & Chowdhury, 2019).

As mentioned above, in this thesis, affective content is considered an umbrella term for content that uses valuation words and affective expressions, which can be simply categorised into either positive or negative language (J. J. Gross, 2015). This definition makes no claim about the actual underlying psychological constructs such as attitudes or experienced emotions, and is based merely on the linguistic properties of the words used in the content. For instance, "good" would be considered a positive affect word, while "hate" would be categorized as negative affect. By counting the occurrence of negative and positive words, we can classify affective content. However, it is important to note that this approach excludes affective content in images, as the assessment of affect in images, despite

technological advances, remains challenging and often produces ambiguous results (Gajarla & Gupta, 2015; Y. Wang & Li, 2015).

As such, both the literature I review in this chapter and the empirical findings in the later chapters cannot be generalized beyond text content, as prior research suggests that for example, sharing behaviour differs between text and images (Bossetta, 2018). I return to this issue in the final chapter of this thesis. In the following paragraphs, I will outline how textual affective content is measured, its prevalence on social media, and how users perceive it on these platforms.

Assessment of Affective Content

The definition of affective content in this thesis follows most closely the approach to how affective content is assessed in the NLP literature on dictionary-based sentiment analysis (Hardeniya & Borikar, 2016). This approach aims to computationally calculate the use of affective language by counting the occurrence of affect words, which requires devising a corpus of affective words for which there are two commonly used techniques. The first is manual dictionary creation, as implemented for example by the General Inquirer (Stone, Dunphy, & Smith, 1966) and Linguistic Inquiry Word Count (LIWC, Pennebaker, Francis, & Booth, 2001), and crowd-sourced dictionaries, such as ANEW (Bradley & Lang, 1999), or LabMT (L. Mitchell, Frank, Harris, Dodds, & Danforth, 2013). Manual dictionaries rely on the bottom-up construction of a dictionary, meaning that the creator has to come up with all the words that constitute the specified construct (see Chapter 5 for an example of this procedure). By contrast, crowd-sourced dictionaries employ a top-down methodology that involves a group of raters who assess the extent to which each of a large number of words expresses the targeted construct. For instance, a rater might be asked to evaluate the degree of

positivity associated with the word “birthday,” and the average of all raters’ evaluations would then be used as the weight of this word in the classification dictionary. The benefit of assessing affective content by counting words in a pre-defined dictionary is that there is clarity about why content received a certain rating. However, both dictionary approaches are unable to capture complex figurative rhetoric such as irony (Weitzel, Prati, & Aguiar, 2016).

An alternative method for assessing affective content uses machine learning. Unsupervised machine learning classifies content using unlabelled data and supervised machine learning classifies content using human-annotated data. Topic modelling is an example of unsupervised machine learning, which uses Latent Dirichlet Allocation (LDA) to find clusters of words that make up latent topics in documents (Blei & Lafferty, 2009). For example, this method could help to identify topics that have specific affective meanings, such as texts about specific political events (see Chapter 2), or even a topic that is made up of mostly affective words (see Chapter 3). Supervised machine learning approaches such as Recursive Neural Tensor Networks, which are a type of deep learning model designed to handle structured data such as text, can be used on human-annotated training data to build a classifier for affective content (e.g. Stanford Sentiment Treebank, Socher et al., 2013). As both machine learning approaches categorize texts by learning latent structures based on mathematical algorithms, it can be unclear from their output why certain texts are classified as affective or not (Ito, Tsubouchi, Sakaji, Yamashita, & Izumi, 2020; H. Li, 2018).

Despite the high accuracy of machine learning approaches and their ability to detect more complex constructs such as irony (for an overview of the strength and weaknesses of both approaches, see H. Li, 2018), I decided to use dictionary-based classifications and definitions of affective content in my doctoral research, as this reduced the ambiguity of what

constitutes affective content. Namely, based on this definition, in this thesis, I investigate social media content that uses various levels of positive and negative affective words. As previously mentioned, simply analysing the affective terms used in peoples' content does not allow me to make direct claims about the emotion that a user felt while creating this content or about how viewing this content will make other users feel. In order to evaluate the relationship between the use of affective terms and felt emotions, social media data has to be combined with experimental or survey data. Investigating this relationship using a multi-method approach is one of the central goals of this thesis.

The primary aim is not to distinguish between different discrete emotions such as anger, sadness, or joy, which could be assessed using different sentiment analysis tools (e.g. NRC Word-Emotion Association Lexicon, also known as EmoLex, Mohammad & Turney, 2013), but rather to examine the broader comparison between positive, negative, and neutral affective content online. As highlighted in the next paragraphs, previous research using even these dichotomous categories has yielded mixed findings regarding which kind of affective content is shared more. To gain a more detailed understanding of affective content, I also employ other methods such as topic modelling.

Engagement with Affective Content Online

There are not only multiple ways to assess affective content on social media, but also a wide range of methods for measuring engagement with it (for a review see Goldenberg & J. J. Gross, 2020). The three most common types of engagement are "liking", commenting, or sharing the content. In February 2009, Facebook introduced the like button as a small thumbs-up symbol on their platform (Gerlitz & Helmond, 2013), which was quickly adopted

by other websites. However, the like button is an ambiguous signal for endorsement, as people may like a post to either support the person, the message, or both (Eranti & Lonkila, 2015). Additionally, “liking” might have the lowest activity threshold as it is the least public form of engagement that requires only minimal effort. Since 2015, Facebook diversified its like-button by adding other reactions such as “haha”, “angry”, “surprised” and “sad”, which allow for a more detailed analysis of users’ opinions (Turnbull & Jenkins, 2016).

In contrast to likes, the commenting feature allows users to express a more detailed opinion on the content they have engaged with, which may include agreement or disagreement with someone’s original message. It allows them to share their thoughts and perspectives on the original message, and say whether they agree or disagree with it. Commenting typically takes two forms: a direct reply to someone’s content, or a quoted retweet on Twitter, which enables users to share someone else’s content with their own text at the beginning, either supporting or criticizing the original content. However, analysing commented content can be particularly challenging. Comment threads can evolve dynamically, with replies to the original post or other users’ comments. This complexity makes it difficult to determine whether the commenter endorses or condemns the content as a whole.

Sharing content without further editing is less ambiguous because it indicates the kind of content that users agree with to the extent that they want to pass it on to other users (Barberá et al., 2015; Tufekci, 2014). Recently, Frimer et al. (2023) suggested using the ratio of retweets to likes as a measure of engagement, arguing that certain content (e.g., extreme political content) might receive many retweets but not be genuinely preferred or liked by most people. However, since the exact meaning of “likes” is unclear, I decided against using

them as a measure of approval. Shares are a clearer endorsement of the message, indicating support not just for the producer but also for the content itself.

In this thesis, I focus on assessing engagement using retweets because they not only show endorsement but also demonstrate that the person has engaged with and read the message to an extent that they want others to see it. Measuring sharing can also indicate the prevalence of certain content, as I discuss in the next section.

Prevalence of Affective Content Online

Similar to our offline conversations, we often discuss various topics online using affective terms. Unlike in face-to-face co-present interactions, online content can reach a large audience simply by being shared with followers with a click of the mouse, and we have constant access to more content that we are able to consume (Aral, 2020; Brady et al., 2019; Brady, Jackson, Lindström, & Crockett, 2023). Because of this overabundance of content, social media is an attention economy in which users compete for each other's attention. This is done by producing content that leads to more engagement (Brady, Gantman, & Van Bavel, 2020; Kramer et al., 2014). One way of attracting more attention is by using affective terms (Robertson et al., 2022; Soroka & McAdams, 2015; Zhang & Qu, 2018). Thus, users are incentivised to use more affective terms, leading to a high frequency of affective content online (Brady, McLoughlin, Doan, & Crockett, 2021).

The content created on social media tends to be predominantly positive in affective tone (Kedrowicz, Royal, & Flammer, 2016; Vogel & Rose, 2016), but people also voice their struggles and pains, express outrage towards political and social issues, and get involved in

conflicts of various kinds (Jalonen, 2014; Vermeulen, Vandebosch, & Heirman, 2018). On many social media platforms, only a relatively small fraction of users produce and distribute the vast majority of negative affective content or content about contentious topics such as politics (Frimer et al., 2023; Kumar, Hancock, Thomas, & Durumeric, 2023). One reason for the prevalence of positive affective content is that users might want to present themselves in a positive light or at least do not want to be perceived as being negative individuals (Lin, Tov, & Qiu, 2014). Users who have a large network and a high need for impression management are especially likely to create more positive content (Hogan, 2010; Lin et al., 2014). Another reason why social media content is predominantly positive, is that human language has a general positivity bias such that more positive affect words are available in the first place (Garcia, Garas, & Schweitzer, 2012; Kloumann, Danforth, Harris, Bliss, & Dodds, 2012). Positive words are also used more frequently than other words across different forms of communication (Dodds et al., 2015). However, the relative proportion of positive and negative affective content does fluctuate depending on the day of the week (Monday being the most negative day, Garcia & Rimé, 2019; Guzman, Azócar, & Li, 2014), time of the year (November and December are exceptionally positive due to holidays, Bollen, Pepe, & Mao, 2009), and the occurrence of recent events such as health crises (Nemes & Kiss, 2021), the Olympic games (Gruzd, Doiron, & Mai, 2011), or changes in the local weather (Baylis et al., 2018; Jiwei Li, Wang, & Hovy, 2014). These documented relations between mood-inducing circumstances and expressed affect provide indirect evidence that there is an association between felt and expressed emotions. The proportion of positive and negative affective content online thus allows researchers to gauge the public mood, which can be used to predict behaviour such as gambling (Otto & Eichstaedt, 2018), stock market trading (Bollen & Mao,

2011; Sousa et al., 2019) and political protests (Bahrami, Findik, Bozkaya, & Balcisoy, 2018).

When it comes to sharing other users' affective content, behaviour on social media presents a puzzle. Despite the higher frequency of positive affective content (Ferrara & Yang, 2015a, 2015b), and despite the general desire to experience positive affective content (M. Tamir, 2009), it is negative affective content that is most likely to be shared with others (Bellovary, Young, & Goldenberg, 2021; Brady, Wills, Jost, Tucker, & Van Bavel, 2017; Fine & Hunt, 2021; Schöne, Parkinson, & Goldenberg, 2021). This is especially surprising as social media users claim that they want to see more positive affective content and that this type of content should be more viral. However, at the same time, people are aware that negative affective content is probably more viral on social media (Rathje, Robertson, Brady, & Van Bavel, 2022). Why, then, is it the case that negative affective content is shared more? It is critical to answer this question, because, in certain user networks, shared content may constitute a substantial percentage of all circulated content (D. Chang & Ghim, 2011), and therefore constitutes a significant proportion of the content that users are exposed to online. In other words, the composition of content viewed by users is not just dictated by what their network originally produces, but also by what it shares. Consequently, if a user's network does indeed have a tendency to share more negative affective content, they could end up encountering a higher proportion of negative affective content than originally created. This discrepancy can skew their perception, making the platform and its users appear more negative than they might actually be, which in turn has consequences for their well-being.

One explanation for the tendency to share more negative content that applies across online and offline content is the negativity bias, which is the tendency to attend more to

negative emotions than to positive emotions (Soroka & McAdams, 2015). The notion of negativity bias is grounded in an evolutionary account that sees negative emotions as containing more useful information about the world than positive emotions (Skowronski & Carlston, 1989; Soroka, 2016). While it is likely that negativity bias plays a role in users' increased engagement with negative affective content, taken alone it cannot easily account for the finding that the virality of negative affective content has increased over time (Frimer et al., 2023).

Studies across different cultures consistently show that negative affective content is shared more frequently than positive affective content. Such research investigating why certain affective content is shared more has focused on a wide range of platforms such as news websites (Bösch, Müller, & Schneider, 2018), forums (Gîfu & Cioca, 2014), blogs (Chau & Xu, 2007), and streaming websites (Hajar, 2016) as well as different cultural contexts including Western societies (e.g. United States and United Kingdom) and Eastern societies (e.g. China, Japan and Taiwan, Hsu et al., 2021). The degree of virality can depend on the ideal affect of a given culture, that is the kind of affect that is most valued and most sought after by members of that culture (Hsu et al., 2021). When content deviates from this ideal affect, it tends to be shared more frequently. For example, Western cultures like the United States typically prefer positive content that evokes high-arousal emotions. Consequently, content that goes against this norm, such as high arousal negative content, has a higher likelihood of going viral as it disrupts the expected emotional patterns. Similarly, in Eastern cultures where low arousal positive emotions are the norm, content that presents a sharp contrast, such as low arousal negative content, tends to be shared more frequently due to its noticeable deviation from the preferred emotional state.

Studies that support the claim that negative affective language is shared more have predominantly focused on the contentious topic of politics (Bellovary et al., 2021; Fine & Hunt, 2021; Frimer et al., 2023; Soroka & McAdams, 2015). Most of these studies have focused on political events perceived as negative or controversial by users who comment on them (Brady et al., 2017; J. H. Gross & Johnson, 2016; Lengauer, Esser, & Berganza, 2012; Rathje et al., 2021).

Another potential complicating factor is the sample of users whose tweets have been investigated. While studies that have collected data from public figures have typically found that negativity is shared more (Bellovary et al., 2021; Fine & Hunt, 2021; Rathje et al., 2021), studies that have focused only on ordinary users outside of the political context have often provided evidence that positive content is shared more (Ferrara & Yang, 2015b; Goldenberg & J. J. Gross, 2020). However, this comparison is further complicated by the fact that the research focusing on public figures has either investigated political leaders or news outlets. In order to reconcile the mixed findings in previous research, we need to (1) compare ordinary and public figures in non-political contexts sharing more general content, and (2) examine sharing tendency during political events that are appraised positively as well as negatively by groups of users.

The social network and its structure present another set of factors that can influence online behaviour (Centola, 2010; Hogan, 2010), and consequently alter the relative prevalence of certain kinds of affective content, in terms of both its creation and sharing. For example, the size and density of a person's network influence the amount of positive and negative affective content users create (Fan, Xu, & Zhao, 2016; Lin et al., 2014). But how do people construct their network? Most previous studies have suggested that people tend to

select ties, defined as other users they connect with online, who express similar views to their own (referred to as homophily, Barberá, 2015; McPherson, Smith-Lovin, & Cook, 2001). Such a selection strategy leads to the formation of echo chambers of like-minded people (Petrov & Proncheva, 2018). More recent studies show that people also tend to connect with others who are more extreme in their views than they are themselves (referred to as acrophily, Goldenberg, Abruzzo, et al., 2022), which consequently results in exposure not only to more polarised views but also to more extreme affective content, leading to the formation of echo chambers of increasing intensity.

Even after quantifying the type of content being created and shared within a network, it remains unclear exactly what content users actually encounter online (Goldenberg & J. J. Gross, 2020). Social media platforms are motivated to keep users engaged with the online content that they provide. Therefore, their algorithms are designed to perpetuate user behaviour by exploiting social learning biases (Rathje et al., 2022). Specifically, users tend to attend and engage with PRestigious, Ingroup, Moral, and Emotional information (especially negative affective content) or PRIME information (Brady et al., 2023). Much of the information that users receive is dictated by algorithms following these principles, including national news (Cetina Presuel & Sierra, 2019), updates from our friends (Latzer, Hollnbuchner, Just, & Saurwein, 2016), and peoples' political beliefs (Huszár et al., 2022). Some commentators argue that the advances in algorithms and the impact on social media users have long since outpaced scientific research (Acerbi, 2019, p. 186). Given this lack of transparency regarding the content that people see, there is a need for experimental studies alongside observational analysis to understand how the use of social media and the associated engagement with online affective content influences users.

Consequences of Negative Emotion Expression on Social Media

Considering the number of people regularly exposed to online emotional content, it is important to better understand the consequences of expressing and sharing affective content online. These consequences encompass both the beneficial and harmful effects on individuals who express or view the content, as well as on the collective. In this section, I outline previous findings for both the sender and the perceiver of positive and negative affective content on social media.

Consequences of Sharing and Expressing Affective Content for Senders.

Shortly after experiencing an emotional event, people tend to share both positive and negative emotions, which is generally referred to as social sharing of emotions (SSE, Rimé, 2007; Rimé, Finkenauer, Luminet, Zech, & Philippot, 1998; Rimé, Páez, Basabe, & Martínez, 2010). SSE is not only perceived to be rewarding for the sharing person in itself (D. I. Tamir & Mitchell, 2012), but also can facilitate beneficial social interactions and foster connection with the interaction partner (Derlega, Winstead, Wong, & Greenspan, 1987; Rimé, 2007). Sharing positive emotions or experiences can prolong the pleasant feeling (Mauss et al., 2011; Tugade & Fredrickson, 2007), and allow people to relive the positive event and the feelings that were associated with it (Langston, 1994), suggesting that people capitalize on the sharing of positive experiences (Choi, Oishi, Shin, & Suh, 2019). Even sharing negative emotions can help the sharer to reduce negative feelings about the event due to the social support that it solicits (Gable, Reis, Impett, & Asher, 2004; Vermeulen et al.,

2018), which builds trust with the conversation partner (Brudner, Fareri, Shehata, & Delgado, 2022; Kim, 2009). On a collective level, social sharing on social media might increase collective resilience; for example, after the Paris terrorist attacks of November 2015, social sharing was associated with a subsequent wave of expressions of solidarity (Garcia & Rimé, 2019). Expressing emotions can also encourage more helpful emotion regulation strategies as it avoids emotion suppression and allows the conversation partner to offer reappraisals and consider alternative viewpoints (DeLongis & Holtzman, 2005; Marroquín, 2011; Zaki & Williams, 2013).

However, social sharing might be a double-edged sword, as it is also potentially associated with a wide range of potentially harmful consequences. For example, social sharing may sometimes involve co-rumination, the excessive and repetitive sharing of problems and dwelling on one's negative feelings together with another person (Parkinson & Simons, 2012; Rose, 2002; Rose, Carlson, & Waller, 2007). Co-rumination potentially offers a trade-off between positive and negative effects, as it increases friendship quality with a rumination partner while also encouraging the internalization of shared problems (Rose, 2021; Spendelow, Simonds, & Avery, 2017). Similar to how sharing positive emotions allows people to relive positive events and emotions, sharing negative emotions can reactivate negative emotions, potentially diminishing one's well-being (Haggard, Robert, & Rose, 2011). Another example demonstrating this potential trade-off is the study mentioned earlier that examined the aftermath of the 2004 Madrid bombings. In addition to the aforementioned positive effects of social sharing reported by Rimé et al. (2010), the extent to which people shared their experience in the week after the event with others was also associated with higher emotional arousal in relation to the terrorist attacks and higher levels of mental rumination three and eight weeks after the event.

How positive and negative emotions and consequences change continuously over time after an event occurred has thus far not been explored. The previous findings are also limited to social sharing as a result of single temporary events, and studies exploring prolonged events are yet to be conducted. In Chapter 4, I address this research gap by investigating both the positive and negative consequences of sharing affective content in response to the ongoing COVID-19 pandemic.

Consequences of Consuming Affective Content for Perceivers.

The affective content we consume shapes the way we feel and think about the world; therefore, when our perception of this affective content deviates from its objective value, this misperception could lead to potentially adverse consequences. It has been shown that even a single presentation of affective images influences peoples' self-reported affect as well as physiological response (Bradley, Miccoli, Escrig, & Lang, 2008). On the other hand, the consumption of affective content can lead to a bias in how we think about events and other people. For example, people who habitually watch thrillers tend to overestimate the prevalence of crime in the real world (Gerbner, J. J. Gross, Morgan, Signorielli, & Shanahan, 2002). To understand the potential consequences of consuming affective content, we first need to ask the question of whether people can accurately estimate the intensity of affective content. There are two perspectives that need to be considered in answering this question, the first perspective pertaining to the transmission of affect between sender and perceiver, and the second pertaining to the aggregation of affective sequences for a given perceiver.

First, the development of NLP sentiment tools is based on the premise that people can successfully encode and characterize affective content because human annotation is used to measure the quality of the developed tools. Indeed, people tend to have high levels of agreement when categorizing affective words and when rating the intensity of affective words and sentences (Hsu et al., 2021; Pennebaker et al., 2001). However, it is unclear whether these judgments correctly assess the intensity intended by the sender of the content. Empirical evidence from the perception of negative affect in corporate emails suggests that the recipient of a negative affect email demonstrates a negative interpretation bias (Sillars & Zorn, 2021). That is, the recipient interprets the negativity as more intense than neutral observers. In another study, Brady et al. (2022) asked Twitter users to rate the intensity of the outrage and happiness contained in presented tweets. They then compared these readers' ratings of outrage and happiness with authors' own ratings of their texts, finding that readers systematically overestimate intensity, especially if they are heavy social media users. This study indicates that while readers may agree on the perceived intensity of the affective content, it may not reflect the authors' intended intensity.

Second, people on- and offline rarely encounter a single expression of affect; rather, they experience sequences of affective stimuli that they need to encode to make sense of their social interaction (Goldenberg, Schöne, et al., 2022). Sequential presentation of affective stimuli is especially prominent on social media platforms, as they are designed to present content in a sequential manner in users' *news feeds*. The perception of these news feeds may have substantial consequences for users - it may alter their perceptions of events and others' opinions, and shape their own viewpoints, eventually affect their emotional states. The previously mentioned study by Brady et al. (2022) showed that people can overestimate the total intensity of sequences of content that express outrage compared to the individual

average of these expressions. However, they only found this difference for sequences that used content where there was a high discrepancy between the sender and perceiver. The mechanism behind this overestimation remains unclear, and it is unknown whether this finding applies to other affective content that does not involve outrage or happiness. Additional indirect support for the notion that people might tend to overestimate emotions presented in sequences of texts comes from the facial expressions and ensemble coding literature. In an experimental study, Goldenberg, Schöne, et al. (2022) demonstrated that people overestimate the intensity of sequentially presented facial expressions of emotions. This amplification effect became stronger as a greater number of faces were presented in a sequence. Goldenberg et al. (2022) explained their finding by arguing that people are more likely to attend to and remember the stronger emotional expressions from a series and that this leads to an upwardly biased summary representation of intensity. Overestimation of the aggregate value of multiple instances has also been found in the size of circles (Albrecht & Scholl, 2010; Haberman, Harp, & Whitney, 2009), and emotions in monoracial and multiracial crowds of faces (Goldenberg, LaFollette, Huang, Weisz, & Cikara, 2022; Goldenberg, Weisz, Sweeny, Cikara, & J. J. Gross, 2020). Despite the importance of understanding the aggregation process for affective content presented in sequences and previous research hinting at a potential overestimation tendency, to date, no experimental research has directly examined this question.

The perception of others' affective content can impact an individual's own affective state as instantiated by the convergence of expressed affect among users when individuals are exposed to content from other users. Although this influence has only been measured using by changes in users' affective expressions, research on computer-mediated communication has provided some evidence that exposure to affective content does indeed influence users'

reported affective state. In one study, participants were asked to text another person using online chat after a negative emotion induction. The results showed convergence between the two participants' self-reported emotions following the exchange (Hancock, Gee, Ciaccio, & Lin, 2008). However, there is little direct evidence that convergence in expressions mirrors a synchronization in subjective affect on social media.

The process whereby emotion transfers to other people in the online domain is usually described as "emotion contagion" (Goldenberg & J. J. Gross, 2020), a phrase that seems to imply that people "catch" other people's emotions as a result of simple (remote) social contact regardless of what caused those emotions or what they are about (for an overview, see Parkinson, 2020). Similarly, most previous research focuses on the specifically emotional content of messages without factoring in the context. However, an emotional message can also influence other people's emotional responses because it informs them about emotion-provoking events to which they then separately react, or because it guides their interpretation and evaluation of those events in a way that encourages a similar emotional orientation to them (social appraisal, Manstead & Fischer, 2001). Thus, it is not only the fact that a message conveys anger that induces a convergent emotional response, but also the fact that it also contains information about annoying news and encourages readers to find that news annoying. Even the random positive and negative events that were selectively shared with unwitting participants in Kramer et al.'s (2017) Facebook experiment (described above) are likely to have played a role in producing the "contagious" effects that the investigators reported. Because the processes underlying the online spread of emotions do not all seem to involve contagion per se, some researchers have favoured the use of more neutral and general terms such as emotional cascades or online emotion convergence which make no specific claims about underlying mechanisms (Alvarez, Garcia, Moreno, & Schweitzer, 2015).

Besides the affective consequence of emotional transfer, there are also behavioural consequences that are associated with exposure to online emotional content. For example, Kramer and colleagues (2014) report a so-called “Withdraw Effect”, which means that people who were exposed to a lower number of affective texts (either positive or negative) were less engaged on Facebook as indicated by the decreased number and length of posts.

As social media platforms are motivated to keep their users on their platforms, their algorithms are designed to cater to users’ consumption tendencies and preferences. Considering the previously highlighted human tendency towards negativity, this might cause harm to users further downstream (Brady et al., 2023; Kelly & Sharot, 2023). For example, due to the negativity bias, users are inclined to attend to and share such negative affective content to a relatively greater extent, which not only leads to them exposing even more people to this type of content, but also to the algorithm consequently showing these users more negative content (Feezell, Wagner, & Conroy, 2021).

Exposure to negative content not only affects users’ short-term emotional state but also their broader well-being. Generally speaking, well-being is comprised of three components: positive emotions, negative emotions, and life satisfaction (Diener et al., 2017). According to this definition, an increase in the frequency of experiencing negative emotions should lead to a reduction in well-being. Indeed, previous research suggests that negative emotion experiences are associated with negative changes in well-being measures such as life satisfaction, affect, happiness, loneliness, and optimism (Diener et al., 2010; Jose, Lim, & Bryant, 2012; Seidlitz & Diener, 1993). Furthermore, as shown by another study in the context of political news, reduced exposure to negative content can improve people’s well-being (Feinberg, Ford, Thai, Gatchpazian, & Lassetter, 2020). Despite this initial evidence,

much of the existing research on the effects of exposure to negativity on well-being has been done outside of social media. In a study investigating web-searches, Kelly and Sharot (2023) found that people's tendency to visit more websites that contained negative affective information was associated with a decrease in well-being. They further argue that the causal relationship between visiting negative affective websites and feeling negative might be bidirectional. On the flip side, research on well-being and social media has mostly focused on well-being as a function of general consumption rather than specific content type (Kross et al., 2020; Steers, Wickham, & Acitelli, 2014; Tandoc Jr, Ferrucci, & Duffy, 2015; Verduyn et al., 2015), although a recent theoretical article argues that these detrimental effects on well-being are caused by negative emotions that are frequently shared online (Steinert & Dennis, 2022).

It is unclear whether sharing and consuming negative affective content should be uniformly viewed as harmful behaviour. As discussed in the previous section, creating and sharing negative affective content might serve a beneficial emotion-regulation function (Zaki & Williams, 2013), and viewing such content might also have some benefits. In fact, being informed about negative events and news has been important for survival, and this may have contributed to the development of the negativity bias (Lazarus, 2021). Social media provides a platform where important negative events can be reported faster than through traditional news outlets (Fitzpatrick, 2018), as regular people can update others in real-time. This has been especially important for political movements such as #MeToo, BlackLivesMatter, and the Arab Spring, where social media activity was crucial in effecting positive societal change (Manikonda, Beigi, Kambhampati, & Liu, 2018; Mundt, Ross, & Burnett, 2018; Wolfsfeld, Segev, & Sheafer, 2013). On the other hand, exposure to only positive content might not be beneficial either as users who are exposed to others' positive affective content might suffer

from the negative social comparison or the fear-of-missing-out which can increase loneliness and reduce well-being (Lewallen & Behm-Morawitz, 2016; Roberts & David, 2020; Vogel, Rose, Okdie, Eckles, & Franz, 2015; Yang, 2016).

Therefore, by understanding creating and sharing of negative affective content on social media, this thesis also contributes empirically to the growing literature on social media and well-being in general. As emphasized in the previous section, to understand how exposure to affective content influences the perceiver, we need to experimentally examine how such content is perceived. In the next section, I outline the structure of the thesis and link the research questions to the specific chapters.

Overview of the Current Thesis

In this chapter, I have defined affective content on social media and described how it is measured. Furthermore, I have reviewed relevant literature, outlining remaining gaps and consequent research questions that I want to address with four projects (Chapters 2-5). These questions can be divided into two broader themes. First, the studies reported in Chapters 2 and 3, which specifically focus on quantifying the prevalence of positive and negative affective content on social media, investigated the type of content created and shared by its users. In these chapters, I particularly emphasise and analyse the difference in prevalence between negative and positive affective content both in creating and sharing, as negative affective content has previously been identified as most harmful to social media users. Second, in the studies reported in Chapters 4 and 5, I investigated the consequences of affective content on social media by taking the perspective of both the sender (Chapter 4) and the perceiver (Chapter 5).

Chapter 2 reports my investigation into how the amount of affective content created and shared by users differs dependent on the characteristics of those users. Specifically, I compared the amount of negative affective content created by public figures and by ordinary users, as well as the likelihood of their negative affective content being shared by other users. To further understand the mechanism that makes negative affective content widespread on social media, I also investigated user characteristics that make it more likely that their negative content is shared.

Chapter 3 reports two studies, which sought to address the question of whether negative affective content is generally shared more across *all* kinds of political events by comparing levels of sharing in response to political events that are appraised by their supporters as either positive and negative. Using NLP methods, I offer additional insights into the affective content associated with the highest likelihood of sharing, and discuss the potential consequences that these findings might have.

The study reported in Chapter 4 mapped the affective expressions of users during an ongoing crisis to explore the affective consequences of social sharing. More specifically, the study assessed how texting about the situation changes the way users express themselves over time, during and after the COVID-19 crisis. In collaboration with another DPhil student (Gloria WS Ma), we also replicated our findings using an event-based diary study to directly assess peoples' affect instead of indirectly inferring it using their online expressions. However, the results reported in the present thesis are limited to the parts of the project that I was solely responsible for, namely, the social media analysis part.

The studies reported in Chapter 5 investigated how users perceive the average expressed emotion in social media content using an experimental design. Experimental

designs allow for the manipulation of the affective content a person is exposed to and the assessment of the impact of exposure on that person. In Experiment 1, I demonstrated that sequential exposure to affective content is associated with an overestimation of the average affective intensity of the sequence. In Experiment 2, I showed how this overestimation leads to an overestimation not only of the intensity of the affective content of the sequence, but also of the affective intensity of the specific topics that the text sequence was about. In Experiments 3–4, I investigated potential mechanisms for these overestimations by demonstrating that they might be due to preferential memory favouring more emotionally intense content (Experiment 3) as well as the perceived representativeness of more intense content (Experiment 4).

Finally, in the general discussion (Chapter 6), I integrate the findings reported in chapters 2-5 with previous literature and devise a framework to understand how users are exposed to affective content and the consequences of that exposure as well as future directions.

CHAPTER 2. NEGATIVE AFFECTIVE CONTENT IS SHARED MORE ON TWITTER FOR PUBLIC FIGURES THAN FOR ORDINARY USERS

A modified version of this chapter is also reported in a submission to the *Proceedings of the National Academy of Sciences Nexus* journal and can be cited as follows:

Schöne, J., Garcia, D., Parkinson, B., & Goldenberg, A. (in press). Negative expressions are shared more on Twitter for public figures than for ordinary users. *PNAS Nexus*. https://doi.org/10.31234/osf.io/wng5v

Abstract

In this chapter, I present the findings of a study that investigates the seemingly contradictory behaviour of social media users that contribute to the prevalence of affective content on social media, specifically, that social media users tend to create more positive affective content, yet negative affective content is more likely to be shared. To understand why, research has thus far focused on psychological processes associated with the content of tweets. In the current study, I investigate if the content producer influences the extent to which their negative content is shared. More specifically, I focus on a group of users that are central to the diffusion of content on social media – public figures. I found that an increase in negativity was associated with a stronger increase in sharing for public figures compared to ordinary users. This effect was explained by two user characteristics, the number of followers (and strength of ties), and the proportion of political tweets. The results shed light on whose negativity is most viral, allowing future research to develop interventions aimed to mitigate overexposure to negative content. The purpose of this chapter was to investigate the prevalence of affective content by examining the affective content *produced by* and *shared of* certain social media user types. These findings support the claim made in Chapter 1 that original content tends to be more positive. However, contrary to the assumption that negative affective content is generally shared more, the findings reveal that this pattern was primarily observed among public figures. The findings illustrate that public figures contribute more to the exposure of negative affective content compared to ordinary users, as the disproportionate sharing of their negative affective content leads to its inflation in prevalence on social media.

Introduction

In Chapter 1, I defined affective content online and reviewed the literature on its prevalence and consequences. In this chapter, I focus on questions arising from studies that have examined the prevalence. To reiterate, such studies have found that most original content on social media is positive in affective tone (Baylis et al., 2018; Dodds et al., 2015; Ferrara & Yang, 2015b). However, there is a growing realization that negative affective content is shared more than positive affective content (Brady et al., 2017; Rathje et al., 2021; Schöne et al., 2021). Users' increased tendency to share negative affective content inflates exposure to this type of content on social media compared to its true proportion in content production. Overexposure to negative affective content is known to have adverse consequences at the individual level, leading to a reduction in well-being (Diener et al., 2010; Jose et al., 2012; Seidlitz & Diener, 1993). At the collective level too, exposure to negative affective content contributes to group polarization and intergroup conflicts (Brady et al., 2022; de Mello, Cheung, & Inzlicht, 2022). Therefore, it is crucial to understand the roots of sharing negative affective content online and its driving mechanisms.

Beyond the observation that negative affective content is more likely to be shared, previous research on negative affective content sharing has mainly focused on specific content-level features and psychological mechanisms that encourage the sharing of negative affective tweets (Brady et al., 2019; Brady et al., 2020; Rathje et al., 2021). In this chapter, I tackle the question in a complementary way by asking: *whose* negative affective content is more likely to be shared? More specifically, I hypothesize that negative affective content produced by public figures is more likely to be shared than negative affective content produced by ordinary users. I argue that this happens because public figures have distinctive characteristics that make their negative affective content more likely to be shared.

To examine these hypotheses, I first replicated the previous finding showing original content on social media tends to be more positive. Second, I assessed whether the degree of negative affect in the content produced by public figures positively predicts the number of retweets to a greater extent than does the degree of negative affect in content produced by ordinary users. I further tested, if this effect was driven by a certain type of public figures. I then compared the user characteristics of public figures and ordinary users, and assessed which of these user characteristics were associated with an increased likelihood that the negative affective content they generated was shared. Finally, I tested whether the differential effect of negativity on sharing for public figures and ordinary users was mediated by their distinctive user characteristics.

Sharing Negative Affective Content on Social Media

Shared content represents up to 75% of all content that people see on social media (D. Chang & Ghim, 2011). It is therefore important to understand what context is more likely to be shared. Generally speaking, language that contains more affective content is more likely to be shared (Goldenberg & J. J. Gross, 2020), but one central question is whether positive or negative emotional language leads to more sharing. Although some studies have suggested that positive content, such as scientific articles (Milkman & Berger, 2014), Olympic games posts (Gruzd et al., 2011), or news articles via email (Kraft, Krupnikov, Milita, Ryan, & Soroka, 2020), is shared more than negative content, other studies have found that negative content tends to be shared more frequently than positive content in other contexts (Brady et al., 2017; Rathje et al., 2021; Schöne et al., 2021). Notably, the sharing of positivity outside of the political context has been primarily observed in content produced by ordinary users, while negativity sharing has been more commonly observed in content by public figures. The

tendency to share negative content can be found in different cultures and platforms, including Facebook (Rathje et al., 2021), Twitter (Bellovary et al., 2021), and Weibo (Fan et al., 2016).

Previous research has suggested a number of reasons why negative affective content might be shared more than positive affective content. There are two proposed reasons that might be particularly linked to user characteristics that make users negative affective content shared more. First, people have heightened attention to negative affective content, also known as the negativity bias (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Soroka, 2016; Soroka, Fournier, & Nir, 2019), leads to more engagement and sharing (Brady et al., 2020; Zhang & Qu, 2018). However, the impact of the negativity bias seems to be moderated by tie strength (Hansen, Arvidsson, Nielsen, Colleoni, & Etter, 2011), with negative affective content being shared more by users with weaker ties to content producers (Berger & Milkman, 2010; Fan et al., 2016). In other words, negativity bias is more likely to amplify negative affective content shared by relative strangers (Berger & Milkman, 2010; Hansen et al., 2011). Therefore, negative expressions should be more viral for producers with a higher proportion of weak ties with other users.

A second reason why negative affective content is more likely to be shared is that social media users are more engaged with content about politics and intergroup relations, which is also more likely to be negative (S. Peterson, 1981; Ruhrmann, Woelke, Maier, & Diehlmann, 2013; Straughan, 1989). Political content is particularly common and popular on Twitter because users are more engaged with politics than the average US population (Wojcik & Hughs, 2019). Intergroup hostility and reputation considerations lead users who are engaged in political contexts to share content negative affective content more (Goldenberg, Garcia, Halperin, Zaki, et al., 2020; Rathje et al., 2021; Schöne et al., 2021). Therefore, users who write about politics more often are more likely to have their negative

affective content shared. It is important to note, however, that attention to political content and negativity sharing may also be driven by bias in the literature towards political figures and news media. Recent research suggests that despite the fact that that Twitter users seem to be more engaged with politics than the average US population (Wojcik & Hughs, 2019) the majority of users (60%) do not follow any political public figures on Twitter (Wojcieszak, Casas, Yu, Nagler, & Tucker, 2022).

Public Figures and Ordinary Users

Verification status is one of the distinguishing features between users on major social media websites such as Twitter, Facebook, and Instagram. Verified users encompassed a wide range of public figures, including politicians, journalists, celebrities, and athletes. At the time of my data collection, the verification status on Twitter indicated whether a user was a public figure authenticated by the platform or not. The verified status of users changed on November 5th , 2022, after which every user was able to verify their account for $8 a month. Before the transition, verified users made up only a small proportion of all users. For example, Twitter has 229 million active users, of which only 420,300 (0.18%) were verified. Despite their relatively small number, verified users are central to the diffusion of content online (González-Bailón & De Domenico, 2021; H. Wang, Lei, & Xu, 2015).

Public figures have distinctive characteristics on social media which may affect the extent to which their negativity would be associated with sharing. The first characteristic is their *high number of followers* (Cha, Haddadi, Benevenuto, & Gummadi, 2010), which often means that many of these ties are weak ties (Wies, Bleier, & Edeling, 2022). Given that weaker ties are more likely to share negative than positive affective content (Fan et al., 2016; Hansen et al., 2011), negative affective content generated by users who have many followers,

such as public figures, is more likely to be shared compared to other types of content. The second characteristic of public figures on social media is that they produce a relatively *higher proportion of political content*. Public figures not only use social media to promote themselves but also to promote social and political causes (Bossetta, 2018; Furlow, 2011; Ross & Bürger, 2014). Additionally, many verified users on social media are political figures, journalists, or other users who specialize in politics, making them more likely to produce political content. Given that negativity is especially likely to be shared in political content (Bellovary et al., 2021; Rathje et al., 2021; Schöne et al., 2021), negative affective content from users who produce a higher proportion of political content is more likely to be shared.

Previous research has already demonstrated a positive link between negativity and increased content sharing when focusing on public figures such as political leaders or news outlets (Bellovary et al., 2021; Fine & Hunt, 2021; Robertson et al., 2022). However, these previous studies have not compared public figures to ordinary users and have not examined for whom the association between negative language and sharing is stronger. This question seems to be crucial if the ultimate goal is to find ways to reduce negativity sharing on social media. Furthermore, given that research on public figures has mostly focused on news media and political figures, it remains unclear whether the observed relationship between negativity and content diffusion can be generalized beyond this specific subset of public figures.

Research Questions and Goals

The first goal of this study is to quantify the prevalence of negative affective content by measuring how much negative affective content is (1) produced and (2) shared. Specifically, I investigate how much negative, positive, and neutral affective content is produced by public figures compared to ordinary users. The second goal is to assess whether

the effect of content negativity on sharing is relatively greater when that content is produced by verified users than when it is produced by ordinary users, and to determine whether the user characteristics mentioned above can account for this difference in content sharing. To achieve the latter goal, I first tested whether negativity boosts sharing more for content that is produced by public figures than for content produced by ordinary users. As there are various types of public figures who can obtain verification status, I further tested if certain types of public figures were more likely to be differentiated from ordinary users in the association between negative language and sharing. I then checked that public figures have relatively more followers and produce a higher proportion of political tweets, before investigating if these characteristics mediate the differential effect of expressed affect on sharing of content produced by public figures and ordinary users. Analysis scripts and data are available at https://osf.io/xqevy/ .

Study 1 Affective Content Produced by and Shared of Public Figures Compared to Ordinary Users

Method

Participants. I collected the data using the Twitter Application Programming Interface (API). In my analysis, I aimed to compare two types of content producers, public figures and ordinary users. On Twitter at the time of my data collection, public figures had a blue checkmark indicating that their account was verified by Twitter and of public interest, while ordinary users were not verified in such a way up until November 5th of 2022. To collect data from both user types, I first collected a list of users and their account descriptions

before extracting their tweets. Based on previous research, I aimed to collect at least 350,000 tweets per group to detect the effect of emotion expression on sharing (Brady et al., 2017; Schöne et al., 2021). For a one-month observation period, I estimated that I had to sample approximately 7,000 ordinary users (assuming that the median number of tweets is 50, Ma, Schöne, & Parkinson, 2022). Because there is no suitable method to sample random ordinary users directly, I extracted account names from randomly sampled tweets. I used the 1% Spritzer stream, a real-time stream of a random selection of 1% of all tweets, to collect random tweets between January 14th to February 13th, 2019, that were produced by ordinary users as indicated by the absence of verification status. I then obtained user names from the producer of these random tweets. After discarding duplicated users, I collected the profile statistics such as their number of followers of 6,681 users from this list and retrieved the tweets they produced in January 2019 (1,927,684 tweets) as well as descriptive information about these tweets including the number of retweets. I removed all tweets that were not in English and non-original tweets, meaning that I removed retweets that did not contain their own added text, resulting in a final sample of 428,223 tweets produced by 6,678 ordinary users. For public figures, I first retrieved a full list of all public figures from the @verified Twitter account (N = 314,373) and their basic profile statistics. Next, I downloaded the tweets they produced in January 2019. After again filtering the tweets that were not in English and were not original, the sample included 2,246,068 tweets produced by 39,241 public figures. To achieve an equal sample size of users with similar Twitter activity, I used propensity score matching to match ordinary users to public figures based on their tweet count (Stuart, King, Imai, & Ho, 2011). This statistical technique helps address possible confounding factors in observational studies driven by inherent differences in samples. First, I calculated a propensity score for each user, indicating their likelihood of tweeting during the

given month. Then, I employed the nearest neighbor method to match verified and ordinary users with similar propensity scores. This approach ensured that the two groups produced a similar number of tweets during the one-month period, reducing any potential distortions resulting from different tweeting behaviors. Each ordinary user was matched to one public figure with the closest tweet count. After matching, the sample included 6,678 users of each type, with 427,502 tweets from public figures and 428,213 tweets from ordinary users. I also repeated the analysis using the full sample of public figures, finding similar results (see Appendix of Chapter 2).

Evaluation of Affective Content. I used the sentiment analysis tool Valence Aware Dictionary and sEntiment Reasoner (VADER, Gilbert & Hutto, 2014) to estimate the affective content of tweets. VADER was specifically developed for sentiment analysis in social media and is especially suited for short texts such as those posted on Twitter (Ribeiro, Araújo, Gonçalves, Gonçalves, & Benevenuto, 2016). VADER employs a hybrid approach of the methods introduced in Chapter 1 by combining manual selection to find the dictionary words with crowed source techniques. Its development started with a manual curation of words and phrases known for their sentiment polarity, which can be either positive, negative, or neutral. These pre-selected words were then given to crowdsourced participants, who estimated the intensity rating for each lexicon item. For each tweet, VADER generates a continuous sentiment score ranging from -1 (extremely negative) to +1 (extremely positive), along with an overall valence categorization (positive, neutral, negative). A tweet is classified as positive if the sentiment score exceeds 0.2, negative if it falls below -0.2, and neutral if the score falls between the two values. For the statistical analysis, I used the continuous sentiment score. I repeated the analysis using a different sentiment analysis tool

(SentiStrength, Thelwall, Buckley, & Paltoglou, 2012, see Appendix Chapter 2 for more details).

Topic Modelling. To assess the proportion of political content in each user's tweets, I first needed to distinguish political tweets from non-political tweets. I used Latent Dirichlet Allocation topic modelling to identify political content in tweets (LDA; Berger & Packard, 2018; Blei & Lafferty, 2009). LDA clusters text into a pre-defined number of topics representing distinct themes. This enabled me to assess the extent to which each sampled user produced political content. I conducted the topic modeling analysis in RStudio (version 4.0.2) using the "topicsmodel" package (Grün, Hornik, & Grün, 2020).

The specificity/generality of the topics that are identified depends on how many of them are preselected by the investigator. If the investigator decides to examine a small number of topics, topics modeling will use broad brush strokes to divide the content but ignore finer distinctions. By contrast, specifying a large number of topics can result in topics that are too specific for the particular research question. Choosing the number of pre-defined clusters is done to balance the specificity and interpretability of the created topics (Jacobi, Van Atteveldt, & Welbers, 2016). My goal in the topic number selection was to find one general political topic using the smallest number of topics possible to avoid having multiple, more specific political topics.

The meaning of a topic was assessed qualitatively by analyzing the words used most frequently in this topic (J. Chang, Gerrish, Wang, Boyd-Graber, & Blei, 2009). The frequency of a word in a topic is expressed in the β-score ("beta-scores"). After manual exploration of the semantic coherence of the topics, I found that using 5 topics created one topic that seemed to be almost exclusively about politics (as indicated by high β-scores for words such as "Trump," "president," "vote," "government" etc.; see Appendix of Chapter 2

for all efforts and details of the identified topics). After deciding on the number of topics, I derived γ-scores from the LDA analysis, which are percentage estimates of the likelihood that each tweet contained each of the specified topics. Based on this criterion, 25.03% of my sampled tweets were categorized as political tweets, which is similar to previous estimates of the quantity of political content on Twitter (Pew Research Centre, 2022).

User-level Variables. My user-level variables were user type, number of followers, and proportion of political tweets. A user was either categorized as the user type of public figure or ordinary user depending on whether the account was verified or not. Verification status and number of followers were extracted from users' basic account information. However, the distribution of the number of followers was skewed and contained zero values. I, therefore, performed a log-modulus transformation ($y = \log(x+1)$) on this variable before conducting my statistical analysis. The proportion of political tweets was calculated as the number of the user's tweets that were categorized as political by the topic modeling analysis (as described above) divided by their total number of tweets.

For verified accounts, I further classified them into several major categories of verified users, including political figures, journalists, news outlets, entertainment, sports, and organizations, and evaluated the tendency of their negative content to be shared by other users. To do so, I employed three classification approaches in conjunction to evaluate these categories. My first approach was to analyze the most frequent words in users' profile descriptions, in order to identify potential categories and build manually curated dictionaries that describe the words used to classify users into their respective types (see Appendix of Chapter 2 for all categories and their corresponding word parts). Secondly, I matched users based on their identifiers with lists from previous research that had already classified them into specific public figure types. These lists include those created by Barberá (2015) as well

as Rathje et al. (2021) for political figures, and the documentation by Bellovary et al. (2021) for media outlets. Finally, I employed the tool "Demographer" (Z. Wang et al., 2019), which utilizes machine learning and natural language processing techniques to infer whether an account belongs to an individual or organization from multilingual social media data. Using these techniques, I was able to categorize around 46.96% of all verified users into accounting for 55.45% of the tweets in my dataset (see 2.1. Table below for exact breakdown).

Table 2.1. *Breakdown of verified user types*

Category	Users	Tweets	User %	Tweets %
Entertainment	4,281	202,767	9.24	11.24
Journalists	9,304	699,737	31.9	24.42
News Outlets	25	9,349	0.43	0.07
Organisation	2,249	203,063	9.26	5.9
Politics	751	22,892	1.04	1.97
Sport	1,284	78,287	3.57	3.37
Other	20,209	977,358	44.56	53.04

Notes. *The table summarizes the absolute count and percentage of tweets of users and by category, from my total sample of verified users before matching.*

Tweet-level Variable. I used the number of retweets as the main dependent variable. Because the distribution of the number of retweets was skewed and contained a high frequency of zeroes, I performed a log-modulus transformation before statistical analysis.

Results

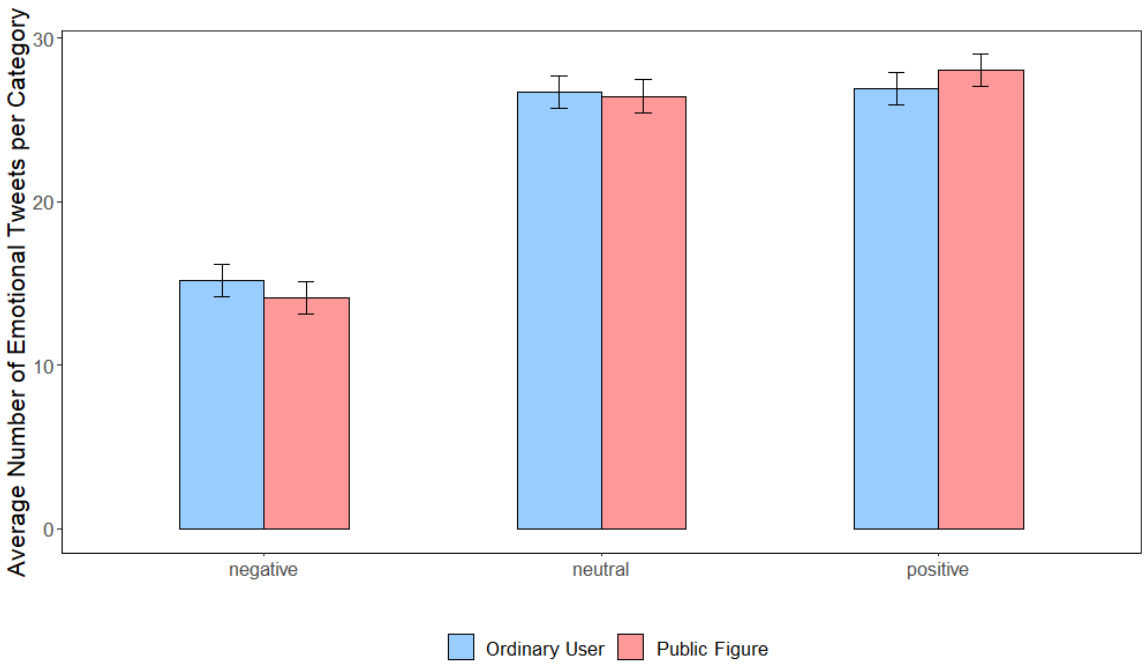
Overview. I first tried to replicate previous research showing that positive compared to neutral and negative affective content is produced more frequently by both public figures and ordinary users. Second, I evaluated if higher levels of negative affective content were more strongly associated with the number of retweets for public figures than for ordinary users. Third, I examined if public figures had more followers, and produced a higher proportion of political tweets than ordinary users. Fourth, I tested if these two user characteristics were associated with increased sharing of negative affective content. Finally, I assessed whether the differences in the user characteristics of public figures and ordinary users could account for differences in the extent to which their negative affective content was shared.

Frequency of Positive, Negative, and Neutral Content for Public Figures and Ordinary Users. I tested if public figures and ordinary users produced more positive compared to negative, neutral affective content using the three VADER categories. For this analysis, I counted the number of a user's tweets in each of VADER's three affective categories. As I matched both user types by their total number of tweets, I was able to compare the absolute number of tweets in the given categories as the dependent variable. Using linear models, I predicted the total number of tweets per affective category based on user type (public figure, and ordinary user).

As expected, positive affective content was more frequent than negative for both user types (b = -12.79 [-13.78, -11.80], SE = 0.50, t (37358) = -25.40, $p < .001$, R^2 = .021, see

Figure 1) [1]. The difference between positive and neutral content was only marginally significant (b = -0.90 [-1.88, 0.083], SE = 0.50, t (37358) = -1.79, p = .072, R^2 = .021). Looking at the comparison between public figures and ordinary users, I found that public figures produced a similar amount of positive content compared to ordinary users (b = 1.09 [-0.30, 2.48], SE = 0.71, t (37358) = 1.53, p = .12, R^2 = .021), a similar amount of neutral content (b = -1.32 [-3.29, 0.64], SE = 1.01, t (37358) = -1.31, p =.18, R^2 = .021), but most importantly less negative content compared to ordinary users (b = -2.15 [-4.12, -0.17], SE = 1.01, t (37358) = -2.13, p = .032, R^2 = .021). These results suggest that public figures tended to produce fewer negative tweets compared to ordinary users.

Figure 2.1. *Number of affective tweets by categories for both user types*



* Marginal R^2 calculated based on recommendations from (Nakagawa, Johnson, & Schielzeth, 2017) using r package "MuMln" (Barton & Barton, 2015)

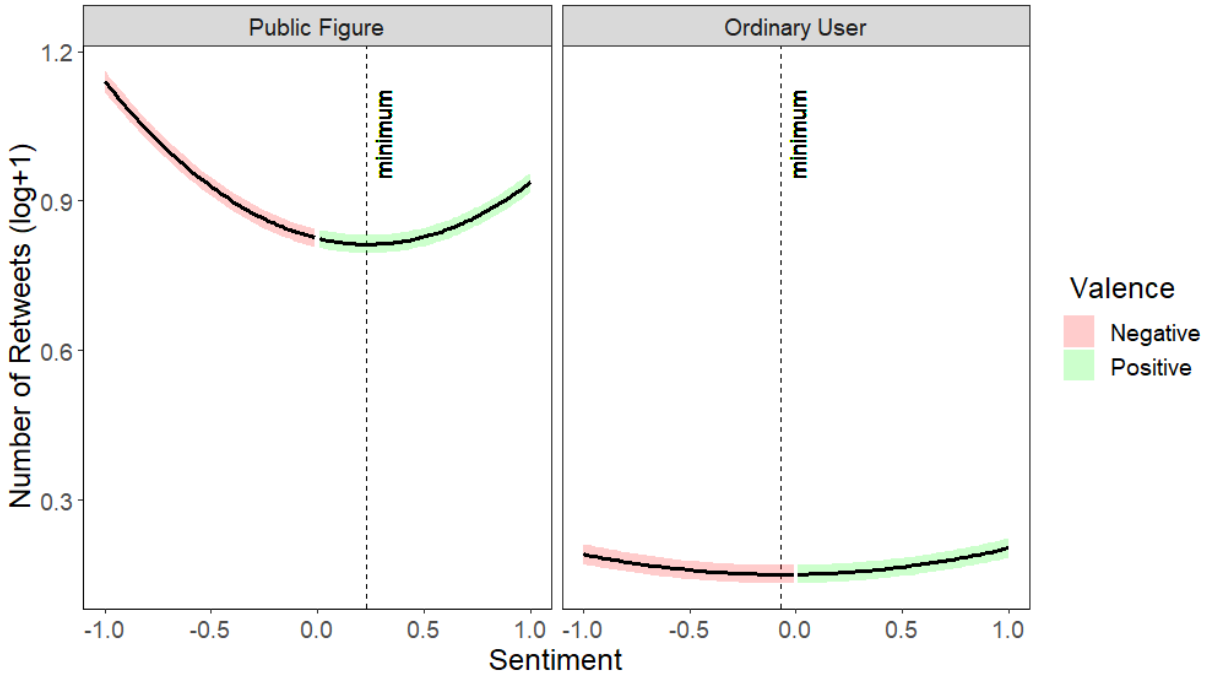
Notes. *The bar-graphs show that for both public figures and ordinary users negative affective content is the least frequent type of originally created content, replicating previous findings. Additionally, public figures seem to produce even less negative content compared to ordinary users.*

Associations between Sentiment Scores and Number of Retweets for Public Figures and Ordinary Users. My dependent variable was content sharing, measured by the log number of retweets. For the predictor, I used the continuous sentiment score from VADER (-1 extremely negative to +1 extremely positive). According to previous research, an increase in either positive or negative affective content should lead to more retweets, therefore I decided to fit a quadratic mixed model to predict the number of retweets using a quadratic function of the continuous sentiment score between -1 and 1. I also investigated potential non-quadratic relationships between sentiment and sharing but the quadratic model seemed to produce the most predictive model (see Appendix of Chapter 2). A quadratic function ($y = ax^2 + bx + c$) returns three coefficients describing the parabola. The coefficient *a* defines how wide the u-shaped graph is and if it opens upwards or downwards. If *a* is positive, then the parabola opens upwards, and if *a* has a higher absolute value, this means that the line slopes more steeply. Coefficient *b* represents whether and to what extent the local peak is a positive or a negative x value. If *b* is negative the local peak is a positive x value and if *b* is positive the local peak is a negative x value. Coefficient *c* is the intercept with the y-axis (at $x = 0$). A larger *a* coefficient in the present model would indicate a stronger association between sentiment and content sharing, while a larger *b* coefficient would suggest that positivity was shared more than negativity. In other words, the coefficient *a* informs us about the overall influence of emotional intensity or extremity on content sharing, while coefficient *b* informs us about whether positivity or negativity is more likely to

be retweeted. I assessed the interaction between user type (public figure and ordinary user) and sentiment in the model and added a random intercept for users to account for baseline differences in the average number of retweets that they received.

Higher sentiment score values were more strongly positively associated with the number of retweets for content produced by public figures than for content produced by ordinary users (a = 0.18 [0.17, 0.19], SE = 0.0067, t(845141.16) = 27.17, $p < .001$, R^2 = .085, see Figure 2.2.). The extent to which an increase in negative sentiment scores led to more content sharing than an increase in positive sentiment scores was also greater for public figures than for ordinary users (b = -0.12 [-0.12, -0.10], SE = 0.0036, t(845324.79) = -31.95, $p < .001$, R^2 = .085).

Figure 2.2. *The number of retweets as a function of sentiment and user type.*



Notes. *The results suggest that stronger sentiment scores are associated with more retweets for both types of users. The local minimum for public figures is reached with a more positive affective tweet,*

indicating that negative affect is more strongly positively associated with the number of retweets for public figures than for ordinary users. Public figures also received more retweets for neutral content than ordinary users.

Having established the fact that the association between negativity and sharing was stronger for public figures, I then examined whether this effect was driven by certain types of public figures. To achieve this, I replicated the previous mixed model analysis, using the log-modulus transformed retweet count as my outcome variable and the interaction of two predictors: the quadratic function of the continuous sentiment score from VADER and the user type category. Unlike my previous analysis, where the user type variable was a simple binary variable (verified users vs. ordinary users), I expanded it to include multiple user types, including ordinary users and seven types of public figures. I also included a random intercept for individual users to account for individual variability.

I found that negativity was shared more for all types of public figures compared to ordinary users, even in comparison to the subset of public figures for whom negativity was associated with the smallest increase in retweets, namely, organizations (b = -0.034 [-0.044, -0.023], SE = 0.0053, t (845324.79) = -6.37, $p < .001$, R^2 = .053, see Table 2.2 for all comparisons). Negativity increased retweets most for political figures (b = -0.32 [-0.35, -0.30], SE = 0.012, t (845324.79) = -26.51, $p < .001$, R^2 = .053), followed by news outlets (b = -0.21 [-0.24, -0.16], SE = 0.020, t (845324.79) = -10.19, $p < .001$, R^2 = .053).

Table 2.2 *Associations between sentiment scores and number of retweets for different public figure types in comparison to ordinary users*

Fixed Effects

	Estimate	*SE*	95% CI	*t*	*p*
Intercept	0.21	0.011	0.19 – 0.23	19.38	<0.001
Sentiment[1]	0.0048	0.003	-0.0011 – 0.011	1.61	0.108
Sentiment[2]	0.049	0.0056	0.038 – 0.060	8.74	<0.001
Entertainment	0.41	0.018	0.38 – 0.45	22.96	<0.001
Journalists	0.44	0.014	0.42 – 0.47	30.66	<0.001
News Outlets	2.48	0.182	2.12 – 2.84	13.61	<0.001
Organization	0.70	0.021	0.65 – 0.74	32.19	<0.001
Others	0.69	0.013	0.66 – 0.71	53.37	<0.001
Politics	0.93	0.029	0.88 – 0.99	31.45	<0.001
Sports	0.46	0.028	0.41 – 0.52	16.60	<0.001
Sentiment[1] × Entertainment	-0.14	0.0055	-0.15 – -0.13	-26.19	<0.001
Sentiment[2] × Entertainment	0.25	0.01	0.23 – 0.27	25.16	<0.001
Sentiment[1] × Journalists	-0.066	0.0038	-0.073 – -0.059	-17.37	<0.001
Sentiment[2] × Journalists	0.098	0.0071	0.084 – 0.11	13.83	<0.001
Sentiment[1] × News Outlets	-0.21	0.020	-0.24 – -0.16	-10.19	<0.001
Sentiment[2] × News Outlets	0.093	0.039	0.016 – 0.17	2.38	0.017
Sentiment[1] × Organization	-0.034	0.0053	-0.045 – -0.023	-6.37	<0.001

Sentiment^2 × Organization	0.043	0.0096	0.024 – 0.062	4.53	<0.001
Sentiment^1 × Others	-0.13	0.0037	-0.14 – -0.12	-36.74	<0.001
Sentiment^2 × Others	0.20	0.0068	0.19 – 0.21	29.92	<0.001
Sentiment^1 × Politics	-0.32	0.012	-0.35 – -0.30	-26.51	<0.001
Sentiment^2 × Politics	0.38	0.022	0.3360 – 0.4241	16.90	<0.001
Sentiment^1 × Sports	-0.10	0.0087	-0.12 – -0.087	-11.95	<0.001
Sentiment^2 × Sports	0.25	0.016	0.22 – 0.28	16.21	<0.001

Random Effects

	Variance	*SD*
User_id (Intercept)	0.90	0.95
Residual	0.73	0.85

Model Fit

R^2	Marginal	Conditional
	0.26	0.56

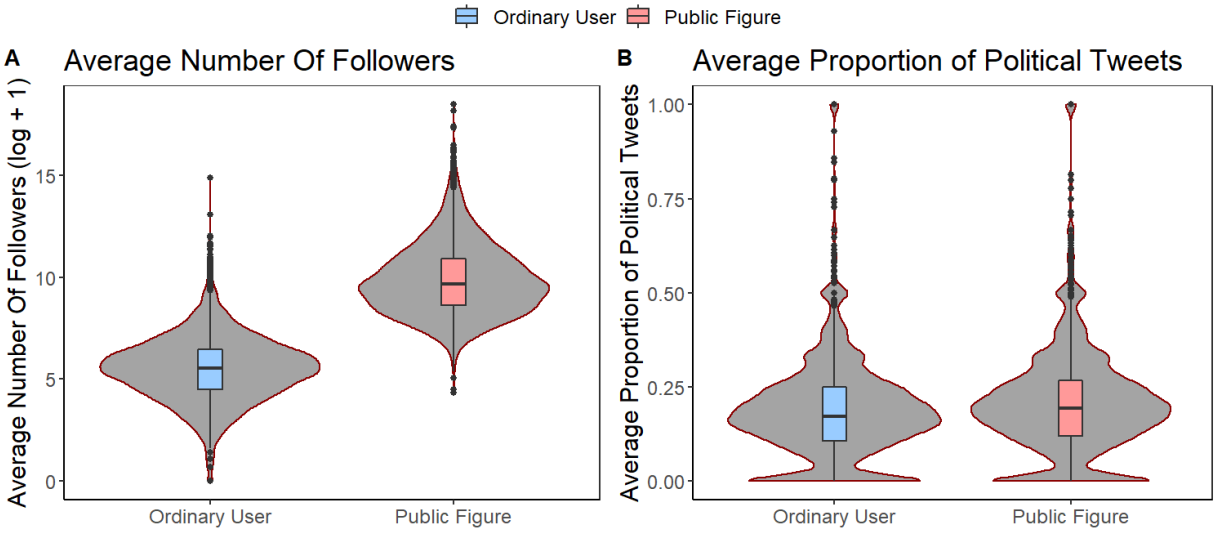
Model equation: log(retweets+1) ~ poly(compound,2, raw = TRUE) * scaled(Number_of_Followers) + (1 | User)

Notes. *A quadratic mixed model with the factors (Sentiment1, Sentiment2, Verified_user_types [7], Sentiment1 * Verified_user_types [7], Sentiment2 * Verified_user_types [7]) and number of retweets (log+1) as the dependent variable.*

Differences between Public Figures and Ordinary Users. I tested if public figures had more followers, and a higher proportion of political tweets than ordinary users using the aggregated user characteristics measures. These were the log-modulus transformed number of the user's followers, and the proportion of their political tweets defined as the number of tweets identified as political tweets by the topic modeling divided by the number of total tweets. I used a simple linear regression predicting the three user characteristics with a dummy coded variable for verification status.

Public figures had a significantly larger following, with an average of 267,258.6 followers (Median = 36,477), compared to ordinary users who had an average of 3,220.60 followers (Median = 449, b = 4.40 [4.34, 4.46], SE = 0.029, $t(13352) = 151.3$, $p < .001$, R^2 = .63, see Figure 2.3.A). Additionally, public figures produced approximately 2% more political content than ordinary users (b = 0.019 [0.014, 0.024], SE = 0.0025, $t(13352) = 7.48$, $p < .001$, R^2 = .041, see Figure 2.3B).

Figure 2.3. *Differences between public figures and ordinary users.*

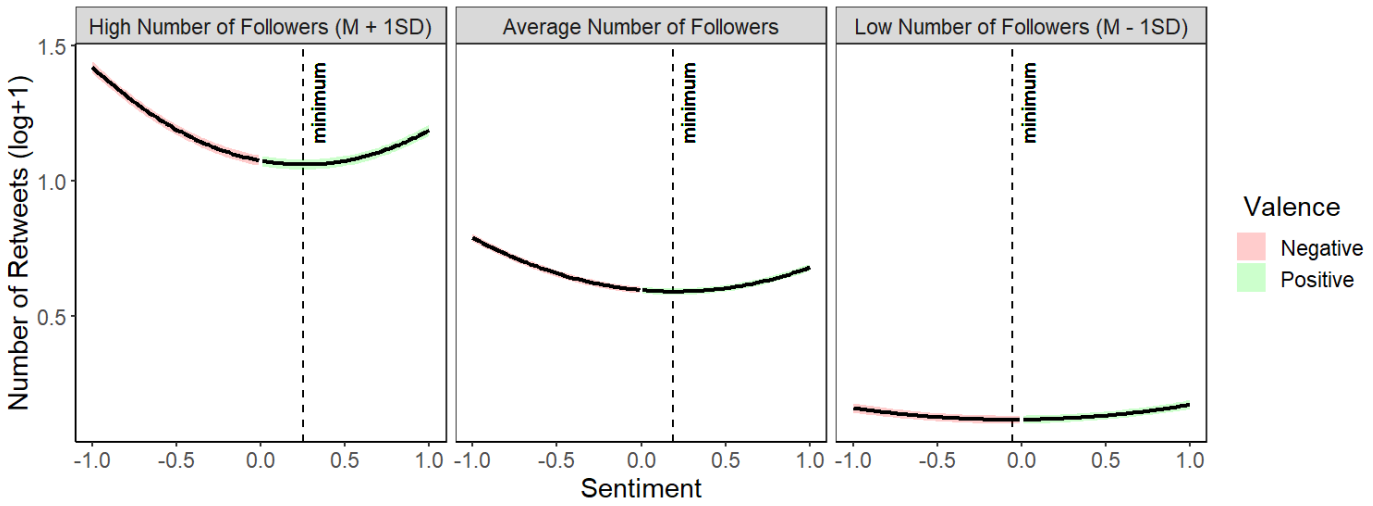


Notes. *Results suggest that public figures have more followers than ordinary users (Panel A) and produce a higher proportion of political tweets (Panel B).*

Association between User Characteristics and Negative Affect Sharing. After confirming that public figures had more followers, and tweeted more about political topics, I tested if these characteristics moderated the effect of affect on the number of retweets. I again used a quadratic mixed model to predict the number of retweets using a quadratic function of the continuous sentiment score. I included each user characteristic as an interaction factor with the sentiment score in two models separately and added a random intercept for users to account for baseline differences in the average number of retweets that they received.

In the first model, I assessed the effect of the interaction between the number of followers (log+1 transformed) and sentiment score on retweets. There was a stronger association between sentiment score and the number of retweets (a = 0.090 [0.083, 0.096], SE = 0.0033, t(846247.48) = 26.88, $p < .001$, R^2 = .21, see Figure 2.4), and between negative affect and number of retweets (b = -0.061 [-0.064, - -0.057], SE = 0.0018, t(846530.08) = -33.29, $p < .001$, R^2 = .21) for tweets produced by users with more followers. Users with more followers also received more retweets for neutral affective content (c = 0.48 [0.47, 0.49], SE = 0.0059, t(12886.53) = 79.97, $p < .001$, R^2 = .21).

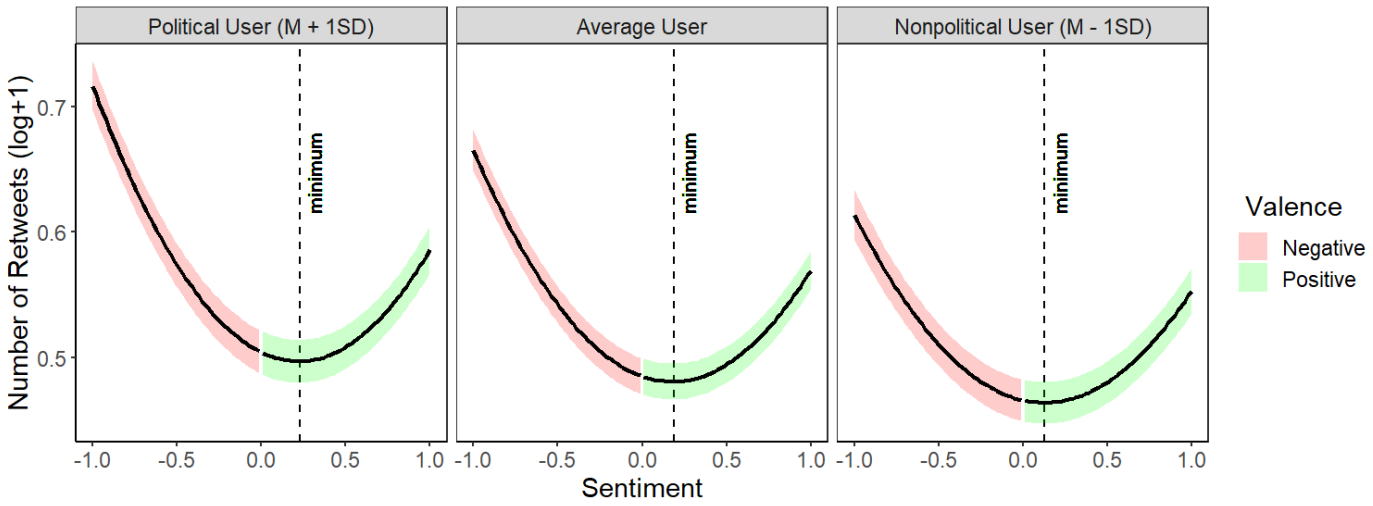
Figure 2.4. *Number of retweets as a function of sentiment and number of followers*



Notes. *To visualize the interaction between two continuous variables (sentiment score and the number of followers), the panels show the mean number of followers in the middle panel as well as the M ± 1 SD (left and right panel respectively). The colored areas on the graph represent the 95% confidence intervals, which are notably narrow. The results show that the effect of sentiment score on content sharing is greater when there is a higher number of followers. In other words, negative affect sharing was stronger for the content of users with more followers.*

In the second model, I examined the effect of the interaction between the proportion of political tweets and sentiment score on the number of retweets. Tweets produced by users with a higher proportion of political tweets showed a stronger association between sentiment score and the number of retweets (a = 0.014 [0.0078, 0.021], SE = 0.0034, t(854258.36) = 4.23, p < .001, R^2 = .014, see Figure 6), and between negative affect and the number of retweets (b = -0.017 [-0.021, -0.013], SE = 0.0019, t(854302.00) = -9.17, p < .001, R^2 = .014).

Figure 2.5 *The number of retweets as a function of emotionality and proportion of political tweets.*



Notes. *To visualize the interaction of two continuous variables (sentiment score and proportion of political tweets), the panels show the predicted association between sentiment score and the number of retweets at the mean proportion of political tweets in the middle panel as well as the M ±1 SD (left and right panel respectively). The colored areas indicate the 95% Confidence intervals. The results indicate that tweets produced by users with a higher proportion of political tweets show stronger associations between sentiment score and content sharing and between negativity and content sharing.*

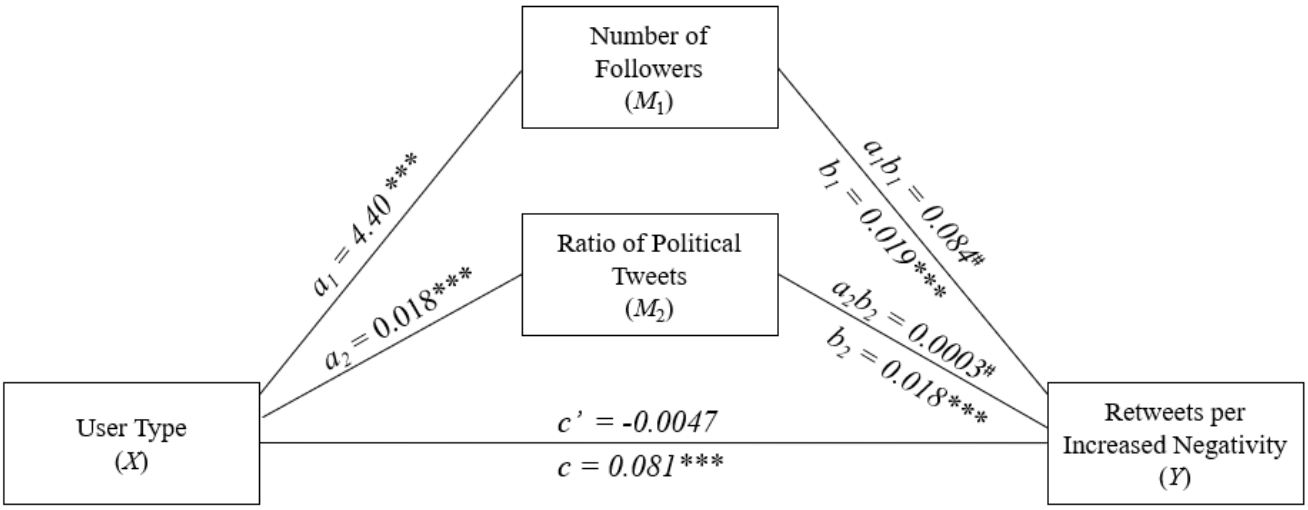
Parallel Mediation Analysis. I hypothesized that negative affective content produced by public figures would be shared more due to the differences in user characteristics that promote negative affect sharing. To assess this prediction, I conducted a parallel mediation analysis assessing two potential mediators of the effects of user type (public figures vs. ordinary users) on sharing of their negative affective content (Hayes, 2017). I computed a new dependent variable that quantified the extent to which an increase in negative affect was associated with more retweets for every individual user in my dataset, which allowed me to

predict how much negative affect is shared for certain users depending on their characteristics. This negative affect sharing dependent variable was calculated using a similar model to the previous ones with two additional changes. First, instead of a quadratic mixed model, I used a split regression to approximate the U-shaped relationship between sentiment and the number of retweets. A split regression contains a categorical variable that is inserted into a linear regression model as an interaction factor to allow for separate slopes for different categories. In my case, I split the continuous sentiment score variable using a binary categorical variable into values < 0 (negative slope) and ≥ 0 (positive slope). This approach allowed me to derive a single coefficient specifically representing the association between an increase in negative affect and the number of retweets. Second, I introduced random slopes representing the relationship between sentiment score and the number of retweets for each user. I extracted the random slope describing the extent to which an increase in negative affect was associated with the number of retweets as my dependent variable for the mediation analysis. Another benefit of using a split regression for the extraction of a per person coefficient was that a linear regression uses one coefficient to describe the relationship between sentiment score and retweets (the beta coefficient), while a quadratic regression uses two (coefficients *a* and *b*, describing how much the sentiment score was associated with retweets as well as if negative affective content was shared more than positive affective content). The potential mediators were the two user characteristics identified above, namely the number of followers, and the proportion of political tweets. I used the PROCESS v4 macro for RStudio by Hayes (2017) to conduct the parallel mediation analysis.

Starting with the a-paths in our parallel mediation model, user type was, as reported above, a significant positive predictor of both number of followers ($a_1 = 4.40$ [4.33, 4.45], $SE = 0.029$, $t\ (13352) = 150.82$, $p < .001$, $R^2 = .63$, see Figure 2.6), and the proportion of

political tweets (a_2 = 0.019 [0.014, 0.024], *SE* = 0.0025, *t* (13352) = 7.47, *p* < .001, R^2 = .0042). Both the number of followers (b_1 = 0.019 [0.018 - 0.0021], *SE* = 0.0006, *t* (13352) = 31.01, *p* <.001, R^2 = 0.13), and the proportion of political tweets were also significant predictors of sharing of negative content (b_2 = 0.018 [0.043 - 0.0032], *SE* = 0.0072, *t* (13352) = 2.55, *p* =.01, R^2 = 0.13). While the total effect of user type on negativity sharing was significant (*c* = 0.081 [0.076, 0.085], *SE* = 0.0022, *t* (13352) = 37.08, *p* <.001, R^2 = 0.13), the direct effect controlling for the mediators rendered this effect nonsignificant (*c*' = -0.0047 [-0.011 - 0.0021], *SE* = 0.0035, *t* (13352) = -1.36, *p* =.17, R^2 = 0.13). A 95% bias-corrected confidence interval based on 10,000 bootstrap samples indicated that the sampled indirect effects of user type on sharing of negative tweets via the number of followers (a_1b_1 = 0.083, *SE* = 0.0039), while holding the other mediator constant, were consistently above zero (0.075 - 0.091). These findings indicate a significant positive indirect effect, suggesting that the number of followers mediates the relationship between user type and sharing of negative tweets. Similarly, the sampled indirect effects via the proportion of political tweets were also consistently above zero (a_2b_2 = 0.0004 [0.0001 - 0.0006], *SE* = 0.0010). These results from the parallel mediation analysis suggest that user type only had an indirect effect on negativity sharing. This effect was mediated both by the user's number of followers and the proportion of political tweets.

Figure 2.6 *Parallel mediation analyses the effect of user type on negative affect sharing via two different user characteristics.*



Notes. *User type positively predicted the user characteristics: number of followers, and proportion of political tweets. The indirect effects of user type on negative affect sharing via the number of followers and proportion of political tweets were both significant.*

Discussion

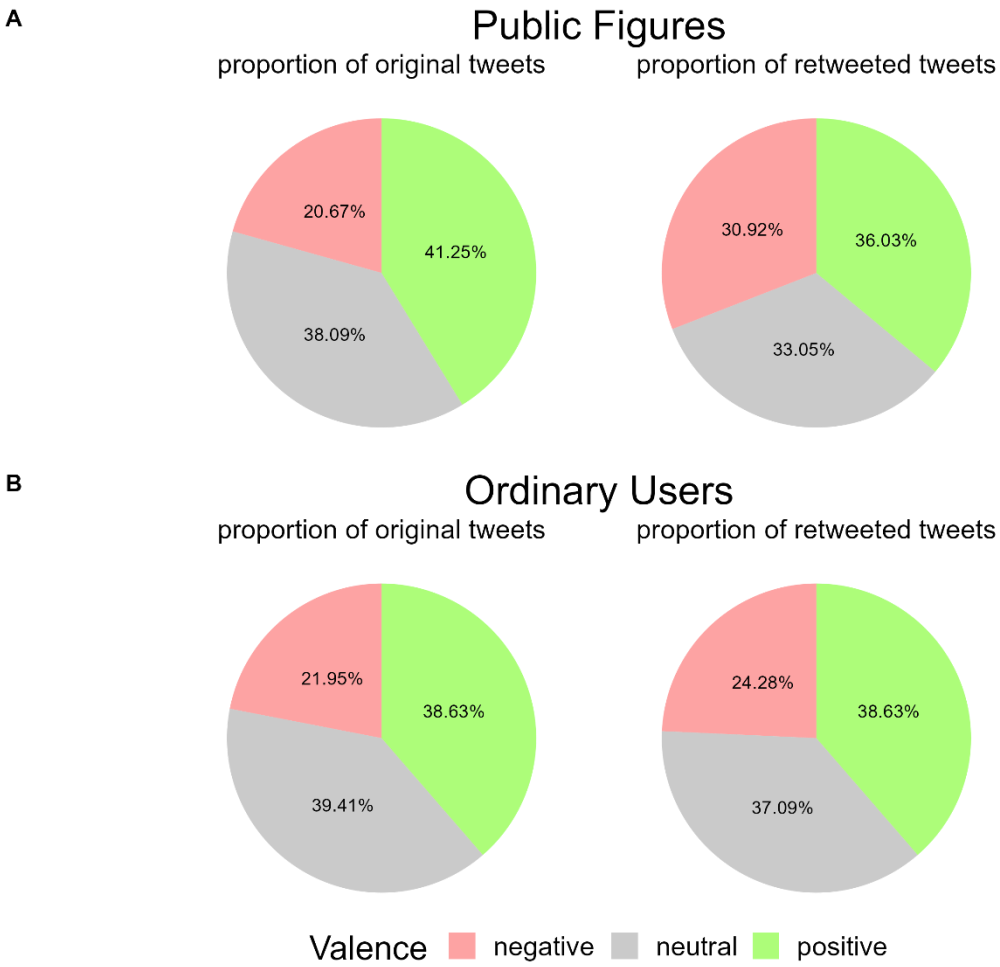
In this study, I compared the extent to which affective content is created by and shared for public figures and ordinary users. I found that higher levels of negative affect in content were more likely to be shared for public users, despite them creating predominantly positive affective content. This stronger association between negativity and sharing was consistent among all types of public figures, while I did not find that negativity was shared more than positivity for ordinary users. I also found that two user characteristics – the number of followers and the proportion of political content – mediated the effect of user type

on the likelihood that their negative affective content gets shared. In particular, the number of followers seemed to be more influential in this relationship as it differed more between the two user types and was a stronger mediator in the mediation model. This work supplements previous research on affective content sharing which has mostly focused on psychological processes elicited by the content features of the tweets, such as the presence of negative affective or morally charged terms. (Bellovary et al., 2021; Brady et al., 2017; Schöne et al., 2021). The current study extends these previous findings by focusing on user-level characteristics that lead to stronger negative affect sharing.

Discussion

Public figures seem to contribute substantially to people's exposure to negative affective content on social media. Interestingly, this study showed that it was not the *amount* of negative affective content they created, but the *degree* to which their negative content was shared, that accounts for their contribution of negative affective content on social media. Whenever a tweet is shared it is duplicated and displayed to the sharer's followers. Considering their large number of followers and how central public figures are in social media networks (González-Bailón & De Domenico, 2021; H. Wang et al., 2015), their content makes up a large share of the material presented on social media. Because negative affective content produced by public figures is shared more, this leads to a disproportionate amount of negative affective content that is unrepresentative of the affective content originally produced by those public figures (see Figure 2.7.). The resulting negatively biased sample of retweeted content then might lead other users to infer that the most credible and popular users on social media platforms use negative affective language, which in turn might negatively influence emotion-expression norms.

Figure 2.7. *Relative frequencies of affective content in original tweets and retweets of public figures (A) compared to ordinary users (B).*



Notes. *Although negative content only made up 20.67% of the original tweets for public figures, it accounted for 30.92% of all retweets, signifying an increase of 10.25%. Conversely, for ordinary users, the proportion of negative content increased only slightly from 21.95% to 24.28% (2.33%). In contrast, when examining content that is less likely to be shared, such as neutral content for both user types, the proportion of such content decreases. Consequently, this content is underrepresented in retweets in comparison to its original frequency. These results suggest that the virality of negative content for public figures can lead to an inflation of their such content, compared to the original texts they produce.*

Disproportionate sharing of public figures' negative affective content has adverse implications for both individuals and collectives on social media. As outlined in Chapter 1, overrepresentation of negative information, such as negative news or online hate, cultivates a more negative or dangerous evaluation of the world (Gerbner, J. J. Gross, Morgan, Signorielli, & Shanahan, 2002), potentially leading to a decrease in social trust (Näsi et al., 2015), and a reduction in subjective well-being (Feinberg et al., 2020; Keipi, Räsänen, Oksanen, Hawdon, & Näsi, 2018). In addition, exposure to negative affective political content has harmful collective consequences such as contributing to group polarization and intergroup conflicts (Brady et al., 2022; de Mello et al., 2022). This study's findings also provide an explanation for why the overrepresentation of negative emotional content produced by public figures has worsened over time (Frimer et al., 2023). The increased sharing of negative affective content incentivizes public figures to generate more of it, thereby perpetuating the cycle of negativity on social media (Brady et al., 2021).

Limitations and Future Directions.

While this study provides new insights about how much negative affective content is created and shared online, and despite my efforts to address alternative hypotheses, the current analysis has limitations that should be addressed in future work. The most important limitation is the observational nature of this study, which means that I could not manipulate user characteristics, while controlling for others that differed between public figures and ordinary users such as average affective content expressed (see Appendix of Chapter 2). In future studies, researchers could examine and manipulate user characteristics, including those that cannot be measured using observational data alone, by using a curated news feed lookalike that allows for the manipulation of such user characteristics.

The second limitation pertains to my assessment of user characteristics. While it is plausible that average tie strength decreases as the number of followers a user has increases, I do not have a direct measure for tie strength, such as reciprocal connections or mutual interactions. In a similar vein, political content was classified using a semi-automated method, namely topic modeling rather than by human annotation or machine learning classifications. This raises the question of how accurately this classification can categorize political content. To alleviate some of these concerns, I tested different configurations of the topic modeling classifications, still finding similar results (see Appendix of Chapter 2). Future work should sample entire networks over time to measure how much users interact with each other to get a more fine-grained measurement of tie strength as well as a user's general tendency to create political content.

The third limitation of my study is that I only collected tweets from a specific time window in January 2019. During that period, two of the largest English-speaking countries on Twitter - the United States and the United Kingdom - were experiencing significant political tensions. The United States was going through an ongoing partial government shutdown that started on December 22, 2018, and lasted throughout January 2019. Meanwhile, the United Kingdom's government voted down another Brexit deal with the European Union. Both events were widely perceived as negative or contentious, and they dominated the political discourse on Twitter during that time. In the topic modelling, I found linguistic traces of these events as they constituted a large portion of the most frequently mentioned political topics in my analysis (see Appendix of Chapter 2). As a result, similar to the previous literature reviewed in Chapter 1, I cannot generalize the finding that negative affective language was associated with more retweets to events that are more celebratory in nature, which I will address in the next Chapter.

In addition to addressing the above-mentioned limitations, future research should seek to develop interventions designed to minimize general overexposure to negative affective content by targeting the above-mentioned user characteristics. I identified users whose negative affective content had the highest tendency to be shared. As negative affective content is in fact produced more rarely than positive affective content (Ferrara & Yang, 2015b), an effective intervention should aim to prevent the followers of such users from disproportionally sharing negative affective content. One possible way of doing this is by educating users about the consequences of sharing negative affective content of users who have a high number of followers or by providing them with feedback about their tendency to share negative affective content produced by much-followed users (van der Meer & Hameleers, 2022).

I believe that my findings emphasize the crucial role that users with large followings such as public figures have in the dissemination of negative affective content. Furthermore, the findings shed light on the mechanisms that are involved in the process of sharing negative affective content and provide the basis for developing interventions aimed to combat the exposure of negative affective content online.

CHAPTER 3. NEGATIVE AFFECTIVE CONTENT IS SHARED MORE THAN POSITIVE AFFECTIVE CONTENT ON TWITTER AFTER BOTH POSITIVE AND NEGATIVE POLITICAL SITUATIONS.

A modified version of this chapter has been published by the Journal of *Affective Science* and can be cited as follows:

Schöne, J. P., Parkinson, B., & Goldenberg, A. (2021). Negativity spreads more than positivity on Twitter after both positive and negative political situations. *Affective Science*, 2, 379-390.

Abstract

What type of affective content is shared more in political discourses on social media? Prior research (including Study 1, reported in the preceding chapter) has mainly concentrated on situations that primarily evoke negative emotions, indicating that negative affective content tends to be shared more frequently. In this context, the current chapter broadens and complements the previous one by examining the sharing of affective content in response to both predominantly positive and negative political situations. In Study 2, I examined affective content sharing in tweets related to the winning and losing parties in the 2016 US elections, finding that increased negativity (but not positivity) predicted content sharing in both situations. In Study 3, I investigated affective content sharing in two separate situations: the celebration of the US Supreme Court's approval of same-sex marriage (positive), and responses to the Ferguson Unrest (negative), finding again that negativity was shared more. These results shed light on the nature of political discourse and engagement. Consistent with the findings of Chapter 2, the study presented in Chapter 3 offers consistent evidence that negative affective content is shared more frequently in both positive and negative political events, highlighting the potentially detrimental effects on individual well-being in this context.

Introduction

In Chapter 1, I reviewed previous research on affective content sharing in political contexts, which has predominantly shown that negative affective content is more likely to be shared than affective content (Brady et al., 2017; Goldenberg & J. J. Gross, 2020). In line with these findings, Study 1 (reported in Chapter 2) found that the negative affective content of users who produced more political tweets was shared more. However, it is important to note that during the period in which I conducted Study 1, the political context was predominantly negative due to events such as the partial government shutdown in the United States and the failed Brexit deal with the European Union in Europe. Similarly, previous research has largely focused on political situations that elicit negative emotions, such as school shootings (Doré, Ort, Braverman, & Ochsner, 2015) or anti-government movements (Alvarez, Garcia, Moreno, & Schweitzer, 2015). Therefore, it remains unclear whether the tendency to share negative affective content more also applies in relation to events that are perceived as positive by users online. Learning about how affective content is shared in response to political celebrations can complement prior findings and provide a more complete view of the nature of political discourse on social media.

One example of such a political event that could provide an opportunity to learn about the affective content sharing in response to more positive situations was the predominantly celebratory response to the U.S. Supreme Court's 5:4 decision that all states under their jurisdiction were required to grant same-sex marriages. This decision was the culmination of a collective action movement directed at allowing gay couples an equal standing before the law. The decision was widely celebrated around the US and the world. But supporters of the movement did not only produce positive affective content to express happiness and joy. Many also shared the frustration they had felt over the years or expressed anger towards those

who opposed them. On Twitter, millions of both positive and negative affective tweets were created and shared online under hashtags such as #loveWins and #SameLove.

One prediction regarding affective content sharing in political celebrations is that an increase in positivity of content leads to more sharing than an increase in negativity. This can be supported by some evidence that after a positive event, people seek out other people with whom they can share their positive emotions (Gable et al., 2004; Isen, 1984). However, these empirical findings did not focus on intergroup situations in which celebrations among ingroup members are often associated with negative experiences among the outgroup (e.g., one group's victory implies another group's defeat). In these competitive intergroup situations, negative emotions directed towards the outgroup are likely to be prevalent and in some cases even exceed positive affect towards the ingroup (Cikara, 2015; Finkel et al., 2020). Furthermore, group members often use negativity to signal one's values and beliefs to other group members (Crockett, 2017; Jordan, Hoffman, Bloom, & Rand, 2016). The goals of the research reported in this chapter are to replicate previous findings derived from studies of negative political events and to empirically assess conflicting predictions about affective content sharing in positive political contexts.

Affective Content Sharing in Response to Political Situations on Social Media

In an era in which much political discourse takes place online (Aral, 2020; Tufekci, 2017), learning more about what type of affective content is shared to a greater extent takes us a step closer to understanding what content lures people into engaging with political content and participating in political discourse. But insights from such analyses are not limited to dynamics on social media, as the psychological constructs that guide online sharing are likely to be similar to those driving offline sharing (see Chapter 6 for a comprehensive

discourse on the applicability of online findings to offline behaviour). Social media provides opportunities to learn about these processes as they naturally unfold on a large scale.

Engagement with certain kinds of linguistic content on social media can be assessed in multiple ways (for an overview, see Goldenberg & J. J. Gross, 2020). A common method is counting how many times a specific piece of content is shared by other users. Unlike other forms of engagement, such as replying or giving a post a "like," sharing content without adding any text is often interpreted as an indication of agreement with the message being conveyed (Goldenberg & J. J. Gross, 2020). Counting shares (or retweets) has the advantage over other measures of engagement in that it provides clear evidence that the user has viewed the original post and is willing to be personally associated with the exact wording and message of the shared content. In contrast, users may simply "like" someone else's content as a show of support, without necessarily endorsing or redistributing the message to their own followers.

It is well established that affective information is shared more frequently than non-affective information, both offline and online (Rimé, 2007; Rimé et al., 1998; Stieglitz & Dang-Xuan, 2013). However, the research focus in relation to specifically political situations has mostly been on events that elicit negative language (Brady et al., 2017; Doré et al., 2015; Garcia & Rimé, 2019). In these contexts, it has been found that negative affective content leads to more shares than positive affective content. However, few political situations are exclusively negative and many of them involve both winners and losers. Understanding affective content sharing in response to political victories can shed light on whether political discourse contributes to intergroup hostility and polarization even in positive situations.

Because positive affective content dominates political celebrations, it may seem plausible that it should also be relatively more likely to be shared than negative affective

content. Congruent with this prediction, outside of social media, Gable and colleagues (2004) found that people seek to share their positive feelings about positive events with others. This is done in order to increase positive social feedback from other people (Langston, 1994). However, available empirical support for this prediction on social media is limited. One relevant study by Brady and colleagues (2017) investigated affective content sharing in tweets relating to the supreme court ruling for same-sex marriage in 2015. Results from their analysis suggested that content was shared more often if it contained words referring to positive moral emotions. However, this study focused on tweets that were written five months after the ruling was made, and the time that had passed since the positive event may have meant that the specific search terms used to generate the dataset (gay rights, gay wedding, or love wins) had already become part of a more general discourse concerning gay relationships and marriage. It is therefore crucial to examine affective content sharing sooner after the target situation has occurred.

By contrast, some accounts provide grounds for predicting that negative affective content should be shared more than positive affective content even in positive political contexts. First, increased attention to negative affective tweets may motivate users to further share these tweets. Previous work has suggested that increased attention to political posts is associated with enhanced engagement (Brady et al., 2020). As negative stimuli are more attention-grabbing than positive stimuli (Meffert, Chung, Joiner, Waks, & Garst, 2006; Rozin & Royzman, 2001), negative political content should similarly encourage engagement. Negative affective expressions may be especially attention-grabbing in intergroup contexts in which negative emotions toward the outgroup are often prevalent and may even exceed the intensity of positive emotions (Cikara, 2015). This is particularly true in the US political context, in which outgroup hate often exceeds ingroup love and is central to group

identification (Finkel et al., 2020). Twitter users may share outgroup-directed negative affective content more because they wish to signal their group membership to other ingroup members and because they believe that ingroup members will relish reading this kind of content (Crockett, 2017; Jordan et al., 2016). The use of negative affective content in an intergroup context may be hedonically pleasing due to the consequent support that users may get from other group members (Cohen-Chen, Pliskin, & Goldenberg, 2020).

Assuming that negative affective content may indeed be shared more in positive political situations, one potential concern is that this is driven by intergroup rather than intragroup interactions in which negative affective content in response to political celebrations is produced by outgroup members. Assessing users' political affiliation is required to evaluate the extent of intragroup versus intergroup interactions.

Research Questions and Goals

The goal of the present research was to examine affective content sharing on Twitter in response to both positive and negative political situations. In Study 2, I tested affective content sharing in response to the US elections in response to tweets that either celebrated Trump's victory or mourned Clinton's loss. Study 3 assessed the generalizability of Study 2's findings by examining data from two separate political situations, the celebration following the supreme court ruling for same-sex marriage and the Ferguson Unrest. Based on the literature, I predicted that content containing negative emotional language would be shared more than content using positive emotional language in situations that are evaluated negatively by users. However, the conflicting theories and evidence did not justify any clear prediction about whether positive or negative affective content is shared more when the

precipitating situation is evaluated positively. Analysis scripts and data are available at https://osf.io/xqevy.

Study 2: Affective Content Sharing for Users with Positive or Negative Evaluations of the Same Situation

In Study 2, I compared the sharing of positive and negative affective tweets that were produced in response to the results of the 2016 US presidential election. I analysed hashtags that were created either to celebrate Donald Trump's victory (positive context) or to mourn Hillary Clinton's loss (negative context).

Method

Participants. I collected data from Twitter using GNIP (gnip.com), which allowed the download of full archives of tweets according to a specified word or hashtag. I focused on tweets of users mourning Hillary Clinton's defeat and users celebrating Donald Trump's victory in the 2016 US presidential elections. I chose to focus on post-election responses because the election result was frequently discussed online by supporters of both parties, who used distinctive hashtags depending on whether their candidate won or lost. I used specific hashtags to identify tweets because previous research suggested that this method should lead to the collection of tweets that are more directly related to the situation and that mostly include users who support the specific cause (Barberá et al., 2015; Tufekci, 2014). The hashtags used to identify tweets mourning Hillary Clinton's election defeat were: #TrumpRiot, #TrumpProtest, #ProtestTrump, #NotMyPresident, #HesNotMyPresident, and #AntiTrump, and the hashtags used to celebrate Donald Trump's election victory were:

#PresidentTrump, #TrumpWinner, #MakeAmericaGreatAgain, #MAGA, #TrumpWon, and #Trump2020. I used the size of the dataset in Brady and colleagues (2017) as a reference point when deciding on the number of tweets required for data collection, setting 100,000 tweets as my minimum data sample with the hope of collecting a greater number of tweets that would permit the detection of even smaller effects. The data collection period was 7 days starting from the date of the election result. I chose this time period to make sure that the hashtags were not utilized to discuss topics other than those directly relating to the specified situation. This process of maintaining the integrity of the hashtags is commonly referred to as "hashtag drift" (Booten, 2016). I collected only original tweets, which are tweets that users created themselves rather than those that included or shared a previous tweet's text such as quoted retweets. I also excluded replies to other users' content to reduce the chance that the emotional language detected was actually in response to another original tweet. By focusing only on original tweets, I avoided additional complexity added by the dynamic of conversation and interactions which cannot be captured by automated tools. For instance, a user may respond with a negative or ironic comment to a positive tweet from someone of a different political party, making it challenging to interpret the meaning of the affective content of the response given this context. To illustrate this concern, consider two instances of positive affective content in response to the same political event: one is an original tweet, such as "Great news! This is a major step forward for equality #LoveWins," while the other is a response of a conservative to a liberal's tweet, like "Wow, what an amazing day for traditional marriage! It's so heart-warming to see the Supreme Court take a bold stand for the sacred institution of marriage. #LoveWins". It would be reasonable to assume that the original tweet is genuinely expressing celebration, whereas the response could be interpreted as an ironic remark directed towards a political opponent. Thus, comparing the motivation

and mechanism behind the sharing of these two types of content would be challenging. I also only collected tweets of non-verified users in order to exclude celebrities, companies, or other organisations whose behaviour has been found to differ significantly from ordinary users (see Study 1, Brady et al., 2017). Finally, I restricted my dataset to tweets that were written in English as the sentiment analysis tools that I employed were limited to the English language. The total number of tweets mourning Hillary Clinton's defeat was 346,730, produced by 161,194 individual users (29.88% of users wrote multiple tweets), and the total number of tweets celebrating Donald Trump's victory was 278,188, produced by 82,970 individual users (28.42% of users wrote multiple tweets). An updated number of shares for each original tweet was collected in a second data collection effort using the Twitter API. As some of these posts were deleted by the users who created them, it is important to note that this sample described above only includes tweets that were still accessible using the API in 2019.

Measures. In contrast to the other empirical chapters, I used the sentiment analysis classifier SentiStrength (Thelwall, Buckley, Paltoglou, Cai, & Kappas, 2010) to evaluate the intensity of positive and negative emotional language used in each tweet. Similar to VADER, SentiStrength is especially suitable for short texts such as tweets due to its capability to integrate syntactic rules such as negation, amplification, and reduction, to interpret the repetition of letters and exclamation points as expressing stronger emotions, and to capture emoticons (Ribeiro et al., 2016). For each individual tweet, SentiStrength provides a score of positive and negative emotional language intensity ranging from 1 (neutral) – 5 (intense). While VADER's sentiment scores combine positive, neutral, and negative scores into a single compound score, SentiStrength's positive and negative scores are independent of each other. This allows us to capture the intensity of positive and negative words used in a tweet separately, which is useful for analysing mixed affective content that uses both positive and

negative language within a single tweet. For example, a text like ‘I love you, but I hate the current political climate!’ would return a positive sentiment score of 3 and a negative sentiment score of 4 in SentiStrength. On the other hand, VADER would return a positive score of 2 and a negative score of 3, after transforming the values onto the same 1-5 scale, because the values of the positive, neutral, and negative categories in VADER always add up to 100%. This means that having high ratings in two categories cannot be accurately represented in VADER’s compound score. Given that my objective in this chapter is to directly compare the increase in retweets resulting from the use of positive and negative language, SentiStrength was the more appropriate tool for sentiment analysis. I subtracted the emotional language intensity scale by one to make it easier to interpret the intercepts of the regression models thus creating 0 - 4 scales. To ensure the consistency of SentiStrength’s results with other sentiment analysis tools, I examined the correlation between SentiStrength and VADER scores in my datasets. The correlation ranged from r = .52 to .62. In addition to the theoretical reason to use SentiStrength for these studies, the model fit using SentiStrength was generally higher, which reinforced my decision to report the SentiStrength results in the main text (see the Appendix of this chapter for a more detailed analysis of the datasets using VADER).

As in the previous chapter, the dependent variable was the number of retweets of the original tweet, which log-modulus transformed before statistical analysis, because the distribution of the number of retweets was skewed and contained a high frequency of zeroes. I do not report how affective content was associated with likes in the main text of this thesis, as likes are often used to signal a variety of reactions other than mere support. However, results from this analysis are reported in the Appendix of Chapter 3.

Results

The Valence of the Situation. My first analysis was designed to confirm that the responses from the two groups of participants sampled using hashtags accurately reflected a positive celebration in one sample and a negative event in the other by examining whether the tweets used more positive or negative emotional language in each respective sample. I used a linear mixed model to compare the intensity of positive and negative emotional language scores for both sets of users. The independent variable was the valence category produced by the sentiment analysis (positive or negative). Our dependent variable was the intensity score corresponding to each valence. As some participants wrote more than one tweet, I added a random intercept of user id to the model to avoid the results being more influenced by certain users who produced more tweets in my sample. For the tweets mourning Hillary Clinton's election loss, the mean score for negative emotional language (M = 0.99, SD = 1.09) was significantly higher than the mean score for positive emotional language (M = 0.43, SD = 0.68, b = -0.56 [0.55, 0.56], SE = 0.002, t (599528.76) = -261.7, $p < .001$, R^2 = .085[2]), verifying that the situation was evaluated as predominantly negative by these users. For tweets celebrating Donald Trump's victory, the mean score for positive emotional language (M = 0.63, SD = 0.77) in the tweets was significantly higher than the mean score for negative emotional language (M = 0.61, SD = 0.94, b = 0.019 [0.015, 0.023], SE = 0.002, t (491557.65) = 8.38, $p < .001$, R^2= .00012) verifying that the situation was evaluated as predominantly positive by these users. However, it is important to note that the difference

[2] Marginal R^2 calculated based on recommendations from Nakagawa et al., 2017) using r package "MuMln" (Barton & Barton, 2015).

between positive and negative scores for Trump supporters was much smaller than for Clinton supporters.

Predicting affective content sharing. My analysis was designed to assess the degree to which positive and negative emotional language scores predicted content sharing. For both situations, I conducted linear mixed model analysis predicting the log-modulus transformed the number of retweets using both the positive and the negative emotional language score for each tweet, as well as their potential interaction. As SentiStrength allows for the possibility of high scores in both the positive and negative categories, an interaction term was included in the model to examine whether the relationship between an increase in one type of valence was particularly strong in relation to the other valence. We again added a random intercept of user id to the model as some users wrote more than one tweet. We also included the user's number of followers as a covariate, because users with more followers generally have more retweets regardless of the emotional content of their tweets as seen in Chapter 2.

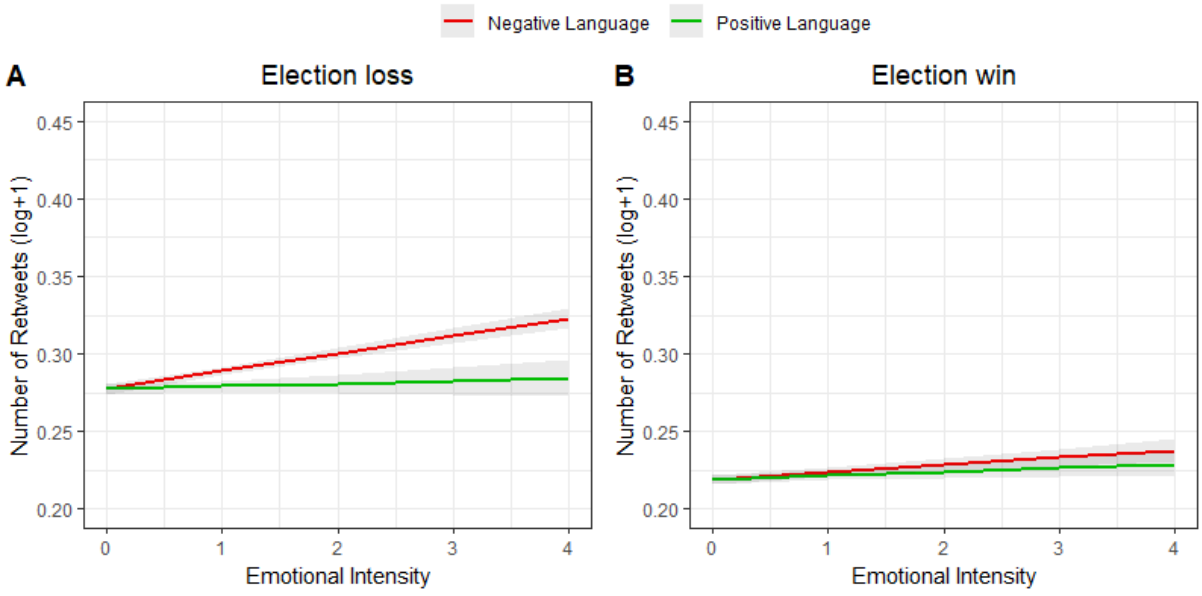
Starting with the results for the tweets mourning Hillary Clinton's election loss, the negative language intensity of tweets positively predicted the number of retweets (b = 0.012 [0.009, 0.014], SE = 0.0013, t (324703.42) = 9.19, p < .001, R^2= .014), whereas the positive language intensity of tweets was not significantly associated with the number of retweets (b = -0.001 [-0.0058, 0.0043], SE = 0.0022, t (321488.52) = -0.48, p = .63, R^2= .014), meaning that when holding the covariates constant, a user received 4.67%[3] more retweets for negative content per increase in emotional language intensity, while the use of positive language did not affect the number of retweets. The interaction between positive and negative emotional

[3] Because the original regression model was conducted on log-modulus transformed number of retweets, the percentages illustrating the size of the effect were calculated by back-transforming the beta coefficients.

language and the number of retweets was not significant (*b* = -0.0014 [-0.0039, 0.0016], *SE* = 0.0015, *t* (318695.10) = -0.93, *p* = .35, R^2= .014).

I then turned to the positive context of tweets celebrating Donald Trump's victory. In this context as well, increased negative language intensity positively predicted the number of retweets (*b* = 0.0045 [0.0014, 0.0078], *SE* = 0.0015, t (263538.25) = 3.16, *p* = .0016, R^2= .024). Specifically, an increase in one unit of intensity was associated with a 1,8% increase in retweets. On the other hand, the intensity of positive language was not significantly associated with the number of retweets (*b* = 0.0018 [-0.00026, 0.0040], *SE* = 0.0010, *t* (267172.38) = 1.72, *p* = .085, R^2= .024, Figure 3.1). The interaction between positive and negative emotional language and the number of retweets was again not significant (*b* = 0.0012 [-0.0019, 0.0041], *SE* = 0.0015, *t* (262868.49) = 0.80, *p* = .42, R^2= .024).

Figure 3.1. *Number of retweets as a function of emotional language used for events evaluated as positive and negative*



Notes. *Results from Study 2 of the degree of emotional language intensity (negative and positive) predicting the number of retweets (log+1) transformed. For both sets of tweets, I found that an*

increase in negative language intensity was associated with an increase in the number of retweets, while positive language was not associated with the number of retweets. For the tweets celebrating Donald Trump's victory (B), I again found that higher negative language scores were associated with a higher number of retweets. Positive language intensity was not associated with the number of retweets.

The Political affiliation of users producing negative affective tweets in the positive situation. The fact that the use of negative emotional language predicted more retweets, even for the tweets that celebrated Trump's victory, called for further analysis of the users' political affiliation. Most importantly, it was necessary to make sure that negative affective tweets in this positive context were mainly written by supporters of the victorious candidate (conservatives), rather than opponents (liberals). If it was mainly Hillary Clinton supporters who used negative language in the tweets that were categorized as celebrating Donald Trump's victory, the increased sharing of these tweets would capture intergroup dynamics rather than intragroup ones. Although it is well established that content produced on social media is mainly driven by intragroup dynamics in echo chambers of like-minded users (Bail et al., 2018; Boutyline & Willer, 2017; Brady et al., 2019), it is possible that the specific context of the election introduced intergroup dynamics too. I assessed this possibility by estimating users' political affiliations.

Previous research has suggested that political affiliation can be predicted by determining which political figures users follow (Bail et al., 2018; Barberá, 2015; Brady et al., 2017; Goldenberg, Garcia, Halperin, Zaki, et al., 2020). Thus, users who follow a greater number of liberal political figures are more likely to be liberals (after excluding extremely popular politicians such as the president or candidates who are running for office from the count, see Barberá, 2015). I used the same approach to estimate users' political affiliation by utilizing Bail et al. (2018) list that contained 4,176 public figures and organizations. To

further understand the sharing of negative language that was written using the hashtags that celebrated Trump's victory, I tested a subsample of posts (30,830 tweets), that were negative (minimum 2 on a scale to 4 in negative intensity), and had at least one retweet. I downloaded all these users' follower lists using the Twitter API and counted how many of the followed accounts were more associated with conservatives (1,602 figures) and liberals (2,574 figures) from Bail et al.'s list. I categorized a user as conservative if they followed more Twitter accounts that were associated with conservatives on the list, and as liberals, if they followed at least as many figures associated with liberals (thus adopting a more statistically conservative measure of political affiliation). Using the method specified above, I was able to estimate the political affiliation of 85.9% of the users I tested. Of the sample of 30,830 tweets 73.44% (22,642) were from conservatives, while the remaining 26.55% (8,188) were from liberals. In other words, my political affiliation classification suggested that the tweets that used negative language in the positive context were produced primarily by conservatives. Following my political affiliation analysis, I repeated my primary analysis of predicting shares from positive and negative emotional language, this time looking only at the subsample of users that I identified as conservatives. This analysis led to the same results as the more inclusive analysis (see Appendix of Chapter 3 for full description).

Analysis of the content of negative tweets. A second question related to the content of the language used in the negative tweets that were produced in the positive context. To investigate the content of these tweets I conducted topic modelling using Latent Dirichlet Allocation (LDA; Berger & Packard, 2018; Blei, 2012). As described in Chapter 2, this method takes texts and estimates latent topics and themes discussed in the texts by measuring word co-occurrence within and across documents (Berger & Packard, 2018). Documents for topic modelling can be a collection of texts such as a chapter of a book,

newspaper articles, or a collection of tweets such as this example. Here, I used this method because it allowed me to (1) assess whether the topics of positive and negative tweets differed from each other, and (2) to qualitatively evaluate the topics that were unique in each of these two groups.

To achieve these two goals, I first classified each tweet as positive or negative on the basis of whether it had a higher positive or negative SentiStrength score (excluding tweets with equal positive and negative scores). I used the "topicsmodel" package to conduct the topic modelling analysis on RStudio (Grün et al., 2020). The number of possible topics was manually pre-defined to two in an attempt to capture the two categories of positive and negative emotional languages and thus maximize the interpretability of the model (Jacobi et al., 2016). These topic models were generated by measuring the co-occurrence of single words (unigrams) and two-word phrases (bigrams).

My analysis focused on tweets celebrating Donald Trump's victory, as I was primarily interested in the topics that were addressed in the positive context. I investigated whether the difference between the topics was based on their calculated valence (positive and negative tweets) by analysing if one topic was predominantly represented in one of the categories and not in the other. A γ-value represents how much a category is associated with a certain topic (per-document topic probability, for an overview, see Silge & Robinson, 2017). In this example a high γ-value means that the words appearing within a category (positive or negative) have a higher probability to only be used in one of the topics, meaning that there are only a few terms that occur often in both categories (Kee, Li, Kong, Tang, & Chuang, 2019). Looking first at the positive tweet subcategory, I found it was mostly associated with the first topic detected by topic modelling. The proportion of words in the positive affective tweet category that are more likely coming from the first topic was γ >

99.9% in the uni- and bigram analyses. The negative affective tweet category was mainly associated with the second topic detected by topic modelling ($\gamma > 99.9\%$) for the uni- and bigram analysis. This meant that results from the sentiment analysis were congruent with the topic modelling analysis such that the topics in the two valenced categories of tweets (positive, negative) were highly distinguishable.

To understand the content of each topic, I examined the unigrams and bigrams that were more likely to be used in one topic but unlikely to be used in the other. The likelihood of a uni- or bigram being used in a specific topic is defined as a beta score β ("beta-scores"). While the words that are most frequent in a topic overlap to a certain extent, comparing words and two-word phrases that are unique to each topic helps to identify the nature of the topic and its underlying content. To do this, I identified the unigrams and bigrams with the greatest differences between the β-values of positive and negative affective tweets using log ratios (Silge & Robinson, 2017): $\log_2(\frac{\beta 1}{\beta 2})$. Table 3.1 shows the top 5 most indicative uni- and bigrams for each topic. The most indicative unigrams and bigrams for the positive affective tweets highlighted the happiness and pride of supporters of the winning party. The hashtag "#TCOT" ("Top Conservatives On Twitter") was the most indicative unigram for the negative affective tweets. TCOT is used by conservatives to locate and identify tweets of other conservative users, which again confirms that these tweets were probably produced by conservatives. The other unigrams seemed to be directed towards opponents such as liberals or the media. The next common distinctive words were "lies" and "hate" which could possibly indicate the troubles that this group had to endure in order to achieve their victory. The most indicative bigram for the negative affective tweets was the conservative phrase "USA Trumptrain", again suggesting that these tweets were created by conservatives. The other bigrams highlighted the still ongoing conflict by talking about the "Trump war" or even

going as far as calling this conflict a second civil war. I have also included the top 10 tweets with the greatest number of retweets from the negative affective category in the Appendix of Chapter 3 to show some examples of these viral negative language tweets.

Table 3.1. *Most indicative retweeted unigrams and bigrams for both positive and negative affective tweet categories from conservatives in response to Donald Trump's election win.*

topic 1 (positive affective tweets):				topic 2 (negative affective tweets)			
Unigrams	$\log_2(\frac{\beta 1}{\beta 2})$	Bigrams	$\log_2(\frac{\beta 1}{\beta 2})$	Unigram	$\log_2(\frac{\beta 1}{\beta 2})$.	Bigrams	$\log_2(\frac{\beta 1}{\beta 2})$.
hope	151.11	wow electionnight	129.53	#tcot	-221.24	USA trumptrain	-134.95
Donald	148.22	Reince MAGA	128.17	liberal	-220.56	war trump	-134.40
love	147.41	sweet patriot	127.68	media	-219.82	beat Hillary	-133.46
happy	147.39	beautiful baby	127.22	lie	-219.32	time ago	-133.21
awesome	147.39	follow god	126.10	hate	-219.26	civil war 2	-133.02

Notes. *This table shows the unigrams and bigrams (bottom table) showing the biggest differences in* $log_2(\frac{\beta 1}{\beta 2})$ *when comparing positive and negative affective tweets ordered by the size of this difference (from top to bottom). Topic 1 shows those with higher* $log_2(\frac{\beta 1}{\beta 2})$ *values for positive affective tweets and Topic 2 shows words with higher absolute* $log_2(\frac{\beta 1}{\beta 2})$ *values for negative affective tweets in the context of celebrating Donald Trump's election victory.*

Discussion

To summarize, Study 2's results indicated that increased use of negative language was associated with a greater number of retweets in both positive and negative situations, whereas increased use of positive language was not associated with the number of retweets. Analysis of users' political affiliation indicated that the negative affective tweets that were produced under the context of celebrating Donald Trump's victory were produced mostly by conservatives. Further analysis of the negative affective texts in this context suggested that the most commonly unique unigrams and bigrams in the negative affective category were used to discuss potential obstacles and adversaries on the path to victory.

Study 3: Affective Content Sharing for Users with Positive or Negative Evaluations of two Separate Situations

One limitation of Study 2 was that it examined the sharing of affective content in the context of the 2016 US presidential election which was one the most contentious political events in the past decade, filled with intergroup conflicts and hostility. A second limitation of Study 2 was that the degree of positivity expressed by Trump supporters was not as strong as the degree of negativity expressed by Clinton supporters. In Study 3, I was hoping to address these limitations by choosing political celebrations that involved less conflict between groups and in which the events were more clearly negative and positive. Additionally, since the users sampled for the positive context in Study 1 were predominantly conservatives, I was hoping to replicate the result by including a political event that was celebrated primarily by liberals. Therefore, in Study 3, I compared two separate political events, one negative and one positive. The negative situation was the Ferguson unrest following the police shooting in

2014 of a black teenager and the positive situation was the 2015 US Supreme Court ruling approving same-sex marriage.

Method

Participants. As in Study 2, I conducted a hashtag search via GNIP to collect data. The negative situation was the unrest in Ferguson in response to the killing of the black teenager Michael Brown on August 9th, 2014, and the positive situation was the US Supreme Court's ruling in favour of Same-Sex Marriage on the 26th of June 2015. My criteria for the selection of these situations were that they were both frequently discussed, distinct (thus removing the possibility of tweets that covered both situations), and specifically identifiable using Twitter hashtags. Again, I only downloaded original tweets (tweets with new content) and did not include replies, retweets, or tweets from verified accounts. Tweets were collected for 7 days following the occurrence of each event. The hashtags used to identify tweets related to the Ferguson unrest were: #Ferguson, #MichaelBrown, #MikeBrown, #Blacklivesmatter, and #raceriotsUSA, and the hashtags used to identify tweets related to the same-sex marriage ruling were: #LoveWins, #LoveisLove, #SameLove, #LGBT, #SCOTUSMarriage, #MarriageEquality. The dataset for the Ferguson unrest event included 552,911 original tweets from 181,179 individual users (39.47% of users wrote multiple tweets), and the dataset for the same-sex marriage ruling included 571,376 tweets from 389,111 individual users (31,89% of users wrote multiple tweets).

Measures

As in Study 2, I used SentiStrength to analyse the tweets' emotional language and counted the number of shares as my dependent variable.

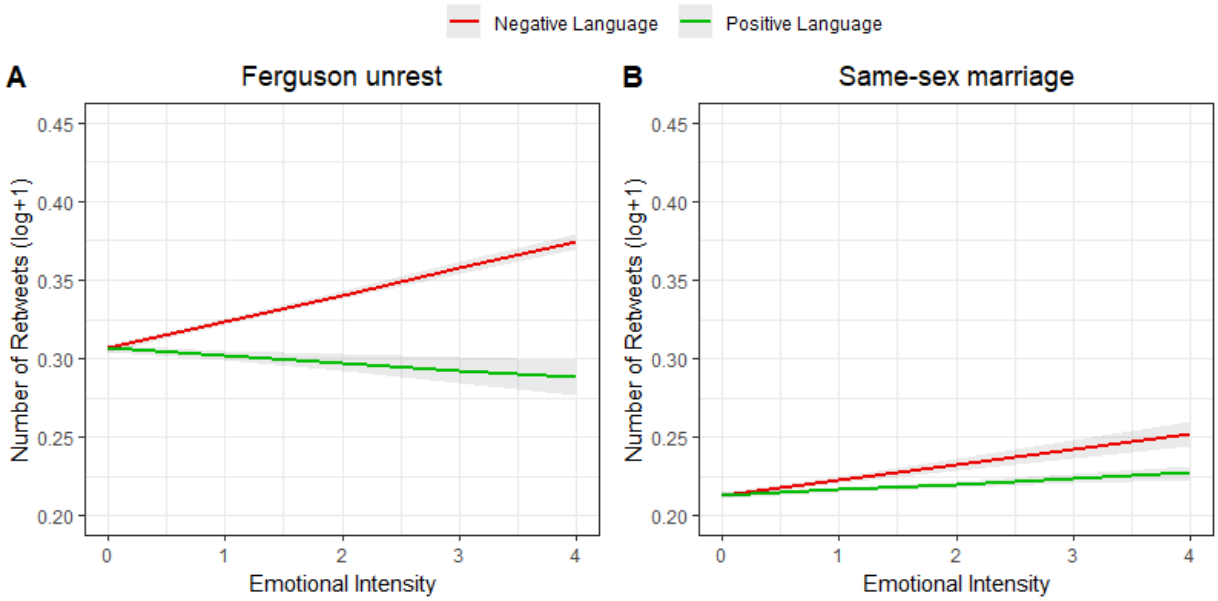
Results

The Valence of the Situation. As in Study 2, I estimated the predominant valence of the emotional language expressed in the tweets by comparing the positive and negative emotional language scores for each situation. Results suggested that the positive emotional language score (M = 0.34, SD = 0.61) was significantly lower than the negative emotional language score for the Ferguson unrest tweets (M = 0.96, SD = 1.05, b = -0.62 [-0.61, -0.63], SE = 0.001, t (1004347.25) = -387.10, $p < .001$, R^2 = .13), whereas the positive emotional language score (M = 0.98, SD = 0.89) was significantly higher for the same-sex marriage tweets (M = 0.37, SD = 0.78, b = 0.60 [0.61, 0.63], SE = 0.0015, t (968932.36) = 390.3, $p <$ .001, R^2 = .12). Notice that unlike in Study 2, the size of the difference between positive and negative emotions in each context was similar indicated by the b coefficients.

Predicting affective content sharing. Using a similar model to the one used in Study 2, I examined the degree to which positive and negative emotional language predicted retweets. Starting with the results for the Ferguson unrest, I found that the negativity of tweets was significantly associated with the number of retweets (b = 0.021 [0.19, 0.22], SE = 0.001, t (541526.13) = 19.08, $p < .001$, R^2= .028), whereas the positivity of tweets was a negative predictor of the number of retweets (b = -0.0081 [-0.011, -0.0031], SE = 0.002, t (538780.62) = -3.94, $p < .001$, R^2= .028). There was no significant interaction between the effects of negativity and positivity on the number of retweets (b = 0.0004 [-0.002, 0.0033], SE = 0.0015, t (538640.47) = 0.33, p = .74, R^2= .028). In other words, an increase in negative language intensity was significantly associated with an increase in retweets, at a rate of approximately 4,77% per intensity unit. On the other hand, the number of retweets decreased by about 1.30% per unit as the intensity of positive language increased.

For the tweets relating to the same-sex marriage ruling, the negativity of tweets was associated with the number of retweets, such that an increase in negativity intensity was associated with a 4.16% increase in retweets (b = -0.0003 [-0.0018, 0.00088], SE = 0.00072, t (943803.74) = -0.50, p = .62, R^2= .051). We also found that for tweets in the positive situation, the positivity of tweets was a significant predictor of the number of retweets, with an increase of 1.5% per unit of intensity (b = 0.0014 [0.00085, 0.0020], SE = 0.00035, t (567799.97) = 4.02, p < .001, R^2 = .051, see Figure 3.2). There was again no significant interaction between the effects of negativity and positivity on the number of retweets (b = -0.0008 [-0.0027, 0.0009], SE = 0.00080, t (881536.00) = -1.04, p = .30, R^2= .051). These results were similar to those of Study 2 with two differences: First, in the negative situation an increase in positive language was associated with a decrease in retweets. Second, positive language significantly predicted an increase in the number of shares in the positive situation, but nonetheless with a smaller coefficient than for the negative language.

Figure 3.2 *Number of retweets as a function of emotional language used for positive and negative events*



Notes. *Results from Study 3 showing how emotional language intensity (negative and positive) predicts the number of retweets (reciprocal+1 transformed). For the Ferguson unrest (A), I found that an increase in negative language intensity was associated with an increase in the number of retweets, while an increase in the positive language was associated with a decrease in retweets. For the same-sex marriage ruling (B), I again found that higher negative language scores were associated with a higher number of retweets. Positive language intensity was also associated with the number of retweets in this context.*

The Political affiliation of users producing negative affective tweets in the positive situation. As in Study 2, I evaluated the political affiliation of the users who wrote negative affective tweets in the positive context (same-sex marriage) to make sure that most of them were indeed liberals. I focused my analysis on users who wrote tweets that used at least moderately negative language (scoring a minimum of 2 on the SentiStrength scale which has a maximum value of 4, N = 4,816). Using the method specified in Study 2, I was able to estimate the political affiliation of 92.11% of the users I tested. Of the sample,

94,45% (4,549) were identified as liberals, while the remaining 5.54% (267) were identified as conservatives. This political affiliation analysis suggested that the negative affective tweets produced in the positive context (same-sex marriage) were primarily written by liberals.

Analysis of the content of negative tweets. As in Study 2, I conducted topic modelling to estimate the content of the tweets that were produced in the positive context, using the same topic modelling method as in Study 2. I again found that there was one topic uniquely associated with the positive affective tweet category and another topic uniquely associated with the negative affective tweet category for both Uni- and Bigram analyses. This distinctiveness of categories was indicated by the γ-scores of > 99.9 each. In other words, the categorisation based on topics modelling was in line with the sentiment analysis, thereby supporting the idea of two distinct categories of positive and negative affective tweets.

See Table 3.2 for 5 unigrams and bigrams that were most indicative of the two categories. The uniquely positive affective words seemed to describe the situation and its significance. The uniquely negative affective words, on the other hand, potentially captured some of the negative connotations of this victory for the supporters of the same-sex marriage ruling. The most common unigram, “tears” is likely expressing tears of joy and signifies positive emotion (see further analysis below to address this issue). Other frequent negative affective words in this category such as “fight” and “hate” were likely used to describe the challenging path that led to the ruling. The most unique negative two-word phrase (bigram) was “ignorant people” which again suggests that these tweets were indeed created by supporters and were used to express negative emotions toward the outgroup. Similarly, the bigram “terrible Christian” confirmed that some of the negative affective tweets were directed at the outgroup. However, I again found combinations including the word “tear”

which could indicate a more complex emotional experience. To make sure that my results are not driven by terms like “tears”, “sobbing” and “cry” which likely represented positive emotions in the present context, I evaluated their contribution to the overall dataset. Looking at the number of words that contained terms related to crying, I found that these words appeared in only 2.7% of tweets. I then removed these tweets from the database and re-ran the above analysis, again finding similar results (see Appendix of this chapter for full description). I have again included the top 10 tweets with the greatest number of retweets from the negative affective tweet category in the Appendix of Chapter 3 to show some examples of these viral negative language tweets.

Table 3.2. *Most indicative retweeted unigrams and bigrams for both positive and negative affective tweet categories from liberals in response to the same-sex marriage ruling*

topic 1 (positive affective tweets):				topic 2 (negative affective tweets)			
Unigrams	$\log_2(\frac{\beta 1}{\beta 2})$	Bigrams	$\log_2(\frac{\beta 1}{\beta 2})$	Unigram	$\log_2(\frac{\beta 1}{\beta 2})$.	Bigrams	$\log_2(\frac{\beta 1}{\beta 2})$.
week	44.31	marriage amaze	48.49	tear	-161	ignorant ppl	-257.25
rule	42.49	magnificent day	46.49	watch	-160.33	tear finally	-254.99
life	41.91	perfect love	46.36	hate	-160.08	terrible Christian	-253.86
decision	41.43	equalmarriage love	45.98	fight	-159.75	marriage dissent	-253.56
nationwide	41.14	beautiful lt	45.69	they’re	-158.94	tear marriageequaility	-253.07

Notes. *This table shows the unigrams and bigrams (bottom table) showing the biggest differences in* $log_2(\frac{\beta 1}{\beta 2})$ *when comparing positive and negative affective tweets ordered by the size of this difference (from top to bottom). Topic 1 shows those with higher* $log_2(\frac{\beta 1}{\beta 2})$ *values for positive affective tweets and Topic 2 shows words with higher absolute* $log_2(\frac{\beta 1}{\beta 2})$ *values for negative affective tweets in the context of celebrating the same-sex marriage ruling.*

Discussion

To summarize, the results of Study 3 replicated those of Study 2, suggesting that the use of negative language was associated with a greater number of retweets in both positive and negative situations. In contrast to Study 2, the use of positive language in Study 3 was associated with an increase in the number of retweets in the context of the positive situation. To alleviate the concern that the sharing of negative language following the positive situation arose from opponents of same-sex marriage, I conducted additional analysis confirming that likely supporters of same-sex marriage in fact produced most of the negative affective content.

General Discussion

The goal of the research reported in this chapter was to investigate the types of emotional language that Twitter users share more in response to both negative and positive political situations. In Studies 2 (same situation) and 3 (separate situations), I examined the sharing of emotional language in response to predominantly negative and positive experiences. The results of my studies indicate that negative language is shared more frequently in both negative and positive situations. Specifically, an increase of one intensity unit of negativity was associated with a 1.8 % to 4.8% increase in retweets, which is in line with previous research that found similar rates ranging between 2% – 6% (Brady et al., 2017; Rathje et al., 2021). Analysis of political affiliation in both studies indicated that the users that produced the negative language in the political celebrations were mainly ingroup

members (conservatives in Study 2 and liberals in Study 3). Analysis of the content of the negative language that was written in response to the celebrations also showed that negative language was mainly used to describe hardships, and past obstacles, or to describe the outgroup in a negative way. Combined, these two studies shed light on the nature of political engagement online.

Implications of the Findings

The fact that negative affective tweets were likely to be shared more even in political celebrations sheds light on the affective nature of political interactions on social media. Although the effect size of negative intensity on the number of retweets seems small, ranging between 1.8 % to 4.8% per intensity unit, it is still of practical importance, given that social media is used by 2 billion users. For instance, let's assume that 10,000 retweets are negative instead of neutral, and using the smallest effect size of the association between negativity and sharing in my sample (1.8%). In this scenario, we could expect an increase in duplication of this content to a total of 180, which will be visible to other users. This is especially troubling because interactions on social media have a reinforcing nature, such that users are likely to produce content that is similar to previous content that generated engagement (Brady et al., 2019; Brady & Crockett, 2019). Increased retweeting of content containing negative language is therefore likely to lead users to further produce more such content and perpetuate the use of negative language (Brady et al., 2021).

Users' preference for negative affective content, even in political celebrations, is in line with the idea that intergroup contexts make negative emotions more prevalent (Cikara, 2015). Assuming that negative language is shared more frequently, and considering the fact that a portion of this language, even in political celebrations, is used to express feelings

toward rival groups, these dynamics are likely to perpetuate more intergroup hostility, which plays a role in affective polarization and sectarianism (Finkel et al., 2020; Iyengar, Lelkes, Levendusky, Malhotra, & Westwood, 2019).

The increase in negative language on social media may also affect perceivers' emotions and general well-being. It is not yet clear whether all forms of engagement with social media are associated with decreased well-being since effects probably vary depending on both the situation and the person (Kross et al., 2020). However, engagement in political contexts, especially when they involve negative language, may be one of the situations in which negative effects are more likely to occur. For example, a longitudinal study demonstrated that daily exposure to political situations was associated with worse physical and mental well-being (Feinberg et al., 2020). Additionally, exposure to hateful content can decrease social trust (Näsi et al., 2015), and exaggerate the feeling of outgroup threat (Lees & Cikara, 2020), which could add to negative well-being outcomes.

Limitations and Future Directions.

Despite the intriguing findings, and the measures I took to dispel alternative hypotheses, there are limitations to Studies 2 and 3 that should be addressed in future research. One possible issue pertains to the question of which psychological states are at the root of the expressions found in negative affective content. In my initial definition presented in Chapter 1, I intentionally refrained from making any assertions regarding the underlying psychological state of the content producer. A key question remains: Does the utilization of negative emotional language accurately reflect negative emotions? One especially important example that highlights this challenge is the word "tear" which was prevalent in the same-sex marriage ruling content and was categorized as negative, while it likely reflected positive

emotions. It is first important to note that removing this word (which represented 2.7% of the tweets) from the analysis did not change the direction or strength of the results. However, it remains important to address the problem of mapping emotional language to emotions. A relevant source of information comes from studies that have evaluated the quality of available sentiment analysis tools and compared them to human annotators. One example is Ribeiro et al.'s (2016) comparison of sentiment analysis tools, which found that in 90% of cases, SentiStrength produced the same emotional category as human annotators. However, larger datasets and improved tools could lead to a more fine-grained analysis of the language used in positive contexts and a better understanding of the content used in these contexts. An alternative approach could involve integrating conventional psychological research techniques, such as self-report surveys, with online data to gain a more comprehensive understanding of the psychological states of users as well as mapping their downstream consequences for well-being. In the next chapter, I will present research that utilizes this approach.

Similarly, a second limitation of these purely observational studies, is that they did not permit me to directly assess the social and psychological mechanisms that encourage or discourage different forms of affective content sharing. In the introduction of this chapter, I provided a few suggestions for potential mechanisms. I believe that the best way to test those mechanisms is to conduct behavioural experiments in which participants see content produced by researchers and are asked to share posts that they like. Tracking users' attention, emotional arousal, and memory for this content may help us further understand the underlying mechanisms. In Chapter 5, I present the findings of a series of experimental studies addressing this research gap.

A third limitation of Studies 2 and 3 arises from the possibility that many of the negative affective tweets that were produced in the positive context were actually parts of intergroup interactions that were not well captured by my analyses. In order to minimize this problem, I focused only on original tweets and removed replies from the dataset. However, it is very hard to evaluate whether a negative affective tweet was written when citing an outgroup news source, or was instead an indirect response to an outgroup user. Future work should use experimental methods to examine the degree to which negativity is elicited even in the absence of any content from outgroup members.

Fourth, there is the question of whether increased negativity is more likely to be shared in political celebrations even in cultures in which expressing intense negativity is less normative and desirable (Heine, Lehman, Markus, & Kitayama, 1999; Tsai, 2019). For instance, studies indicate that Japanese Twitter users tend to produce more low-arousal content than their Western counterparts due to differences in cultural affective values (Hsu et al., 2021). This could suggest that in East Asian countries like Japan, where there is a preference for less intense emotions (Heine et al., 1999), users would also not share intense negativity as it violates their cultural affective norms. One initial answer to this question results from a study by Fan and colleagues (2016), that examined the emotional contagion of positive and negative affective content on Weibo, which is a Chinese social network. Their results showed that anger spread further and faster than joy even in general interactions on Weibo, suggesting that the increased sharing of negativity is cross-cultural. However, this study analysed an unfiltered sample of online content and further studies should examine the sharing of emotions in specifically political contexts across different cultures. These observations raise a broader question as these findings suggests that it is predominantly high-arousal emotions that are shared on social media. Specifically, whether the act of sharing is

influenced by the intensity or arousal of the affective content, or by its valence (positive or negative). In Chapter 6, I will delve into these possibilities in greater depth.

A final question is whether these findings could be generalized outside of social media. Although many researchers have argued that hostile intergroup dynamics are amplified on social media, either due to the nature of social interactions or by top-down algorithms promoting more emotional content (Goldenberg & J. J. Gross, 2020), it is likely that these dynamics represent some general tendencies that people have. For example, although it is common to believe that political echo chambers are especially prevalent on social media, recent work has found that segregation based on physical location may be even greater (Brown & Enos, 2021). Further work should examine emotional dynamics as they unfold outside of social media.

Despite these limitations, I believe that my findings reveal an important aspect of political discourse on social media, addressing the research gap identified in the previous chapter and other research by demonstrating that negative affective content is not only shared in negative political events but also positive ones. My results pose interesting questions and open the door to future investigations and even the possibility of interventions designed to mitigate negative affect sharing. However, thus far, Chapters 2 and 3, as well as previous research, have mainly regarded the prevalence of negative affective content as a negative phenomenon that could potentially harm well-being and contribute to political polarization. As emphasized in Chapter 1, there may be instances in which the expression and sharing of affective content can have positive outcomes for social media users. This concept will be explored in the next chapter.

CHAPTER 4. ONLINE SOCIAL SHARING OF AFFECTIVE CONTENT DURING THE COLLECTIVE CRISIS OF COVID-19

A modified version of this chapter is currently under review at the *British Journal of Psychology*. In the original submission, the project was divided into two parts: a survey portion led by fellow DPhil student Gloria WS Ma, and a social media (Twitter) analysis portion for which I was responsible. This thesis reports only on the portion I led, while also referencing and integrating the findings of the other section in the discussion part. The submitted full version can be cited as follows:

Ma, G. W. S*., Schöne, J.P.*, & Parkinson, B. (2022). Social sharing of emotion during the collective crisis of COVID-19. Available at: https://osf.io/9p3wh/

*Equal Contribution

Abstract

The aim of this chapter is to address two distinct sets of questions. First, in continuation of the research carried out in Chapters 2 and 3, this chapter reports the investigation into the question of what kind of affective content is prevalent on social media. Second, this chapter seeks to explore the affective consequences of sharing and creating affective content for content sender, which has not been addressed previously in this thesis. The primary objective of this chapter is to answer the latter question. To do so, I tracked and analysed the affect expressed in tweets posted over the course of a month in the crisis period. I found that despite expressing more negative and less positive affect following the onset of the COVID-19 pandemic, users on social media still tended to create and share content that was predominantly positive. In contrast to the negative events examined in Chapter 3, users' negative affective content was not more viral, suggesting that negativity might only be viral for ordinary users in the context of politics. Regarding the emotional consequences of sharing and creating COVID-19 content, my research indicated that doing so was associated with an increase in positive affect and a decrease in negative affect related to the pandemic. Overall, the research of this chapter suggested that social sharing is linked with emotional recovery and therefore appears to be an adaptive response to a persistent collective crisis.

Introduction

In Chapters 2 and 3, I analysed the prevalence of affective content on social media by examining the types of affective content that users share and produce. The findings revealed a strong tendency among users to create and share affective content. This raises the question: Why do people express and share affective content on social media and what are the consequences of doing so? As proposed in Chapter 1, one potential explanation for this behaviour is the process of social sharing of emotions (SSE), which suggests that when people experience any emotion, they have a tendency to approach others and discuss it within days, with an 80-95% likelihood (for an overview, see Rimé, 2009). The term "social sharing" within the context of this chapter encapsulates both primary and secondary social sharing. The former involves the generation of original content, while the latter encompasses actions like retweeting another's message. In addition to extending the findings about the prevalence of creating and sharing affective content presented in Chapters 2 and 3 to a different context, the research presented in this chapter focuses on the consequences of creating and sharing affective content online through the lens of SSE, which in turn might shed some light on why users engage in this behaviour.

Sharing and creating affective content on social media might serve a functional role as people might be intrinsically motivated to engage in social sharing as a form of interpersonal emotion regulation (Zaki & Williams, 2013). Positive events may elicit social sharing behaviours aimed at capitalizing on the event, bragging, or mass-sharing (Palmer, Ramsey, Morey, & Gentzler, 2016), thereby prolonging the positive experience or allowing sharers to relive it (Mauss et al., 2011; Tugade & Fredrickson, 2007). In contrast, negative events may encourage individuals to seek social support that potentially improves their well-being (Duprez, Christophe, Rimé, Congard, & Antoine, 2015; Harlow & Cantor, 1995), which is

known as their emotion-recovery goal. However, it is still unclear whether social sharing consistently achieves this goal for individuals in response to negative events, as discussed in detail in Chapter 1. Some studies suggest that social sharing can be beneficial by bringing interactants closer together (Collins & Miller, 1994; Rimé et al., 2010) and allowing sharers to feel socially integrated and experience positive collective feelings such as solidarity and hope (for a review, see De Rivera & Páez, 2007). Other studies, however, propose that social sharing may have negative consequences by reactivating unwanted emotions and thoughts associated with the negative event (Kanyangara, Rimé, Philippot, & Yzerbyt, 2007; Rimé, Bouchat, Paquot, & Giglio, 2020).

To date, only two empirical studies have included measures to evaluate both positive and negative psychological consequences of social sharing (Garcia & Rimé, 2019; Rimé et al., 2010). Doing so is crucial if positive and negative emotions are considered two separate dimensions (Cacioppo & Berntson, 1994). This view is supported by the subjective well-being literature, which suggests that well-being is comprised of independent positive and negative emotional experiences (Diener, Emmons, Larsen, & Griffin, 1985), and diminished well-being can be a consequence of the presence of negativity, the absence of positivity, or both (Iasiello & Van Agteren, 2020; Keyes, 2007). Thus, investigating changes in both positive and negative affect concurrently is important when investigating the consequences of social sharing.

In a study of the aftermath of the 2004 Madrid bombings, Rimé et al. (2010) explored both positive and negative consequences of social sharing and found that the initial social sharing was positively associated with higher mental rumination after three and eight weeks following the event, but it also predicted increased positive affect, reduced loneliness, and stronger perceptions of social support during those weeks. However, this study did not

persistently follow the trajectory of both negative and positive consequences over time; instead, they only measured the affective consequences of SSE by asking people at two discrete time points. Depending on the time points, the presence of beneficial versus harmful effects may differ. For instance, while SSE might alleviate some negative feelings immediately after sharing (for an overview, see Rose, 2021), it might lead to harmful consequences later on as people may ruminate too much about the event (Rose, 2002).

Moreover, all of the studies mentioned thus far have explored social sharing in response to a one-shot negative event. This chapter's research aims to fill these gaps in the literature by exploring online social sharing during an ongoing crisis, the COVID-19 pandemic. As a result of the lockdown measures, many individuals relied on social media to connect with others and share their hopes as well as worries in regard to the ongoing pandemic, providing me with the opportunity to track changes in affective expressions over time and to examine the associated affective consequences of SSE. It is important to acknowledge that merely monitoring online expressions may not fully reflect genuine emotional changes, as previously mentioned (I return to the discussion of this concern in Chapter 6).

Affective Social Sharing on Social Media

Social media has become a natural avenue for communication, and its users have discovered its potential for SSE (Vermeulen et al., 2018). One of the advantages of social media is that users can find an audience without the need for them to be physically present, with just a click of a mouse. For researchers, social media presents equal potential as it is a natural and dynamic platform to observe the SSE. Moreover, using social media to study SSE can overcome one of the biggest limitations of previous SSE studies: the inability to

continuously track the unfolding of both positive and negative consequences of SSE. Additionally, social media allows researchers to measure not only the consequences of SSE, but also the scale and intensity of an individual's engagement in SSE over a specific time span, bypassing the need for potentially unreliable self-reported retrospective estimations.

While people engage in social sharing in response to positive events (Palmer et al., 2016), social sharing is frequently prompted by negative events. Early studies outside of social media by Schoenberg, Carr, Peretz, Kutscher, and Cherico (1975) and G. W. Mitchell and Glicksman (1977) found that approximately 80% of individuals expressed a need to talk in response to acute stressors, such as the death of a loved one or a cancer diagnosis. Similarly, a more recent study conducted after major collective traumas has shown that talking with others helps regulate emotions, such as the 98% of American adults who stated that talking helped them after the September 11 attacks (Schuster et al., 2001). According to the results presented in Study 2 – 3 (in Chapter 3), individuals may have a greater inclination to share emotional content on social media following negative events, as evidenced by the higher average number of retweets for negative events regardless of the type of emotion expressed (as seen in the intercept of both models). This observation underscores the need to further investigate the emotional effects of social media engagement in response to negative events.

A study by Vermeulen et al. (2018) showed that users consider the affective norms of a platform to decide where to share positive and negative events before sharing or expressing any type of affective content. While Facebook, Instagram, and Snapchat are mostly used for sharing positive affective content (if affective content is expressed at all), Twitter and Messenger are also used for sharing negative affective content (Waterloo, Baumgartner, Peter, & Valkenburg, 2018), which makes Twitter an especially suitable platform for

observing social media engagement behaviour in response to negative events. Although previous literature and the discussions in this thesis so far have largely regarded the creation and sharing of negative affective content on social media as detrimental (Rathje et al., 2022; Schöne, Garcia, Parkinson, & Goldenberg, 2022; Schöne et al., 2021), some users have reported that venting about negative events on these platforms provides them with a sense of "relief" (Vermeulen et al., 2018). From an SSE perspective, the sharing of negative affective content can be viewed in a more positive light, but this perspective does not offer a clear consensus on whether it is ultimately beneficial or harmful. As a result, there are competing hypotheses regarding the potential outcomes of sharing either type of affective content.

One previous empirical study by Garcia and Rimé (2019) has directly assessed both the positive and negative consequences of sharing negative content on social media. These investigators focused on online public discourse in the aftermath of the coordinated terrorist attacks that occurred in Paris, France in November 2015, which resulted in 130 deaths and over 400 injuries (Garcia & Rimé, 2019). The findings indicated an increase in French-language tweets generated by individuals residing in France, surpassing double the usual volume on the day of the attacks. Moreover, an observable surge in negative affective content accompanied a significant decline in the employment of positive affective content, hinting at a shared experience of negative emotions. Simultaneously, the use of solidarity-related terms was found to peak one day after the attacks, potentially indicating positive psychosocial effects associated with expressing negative affective content on social media. Despite these initial findings regarding the question of whether social sharing is more beneficial than harmful this study had an array of limitations that afforded further investigation.

One of the limitations of most previous social sharing studies, including the previously mentioned one, is that they have focused on events that are both local in time and

place. For instance, when examining a localized event like the French terror attacks, the findings may yield outcomes that are very specific to this particular context. Comparing the expressions of solidarity following the French attacks to those following the 2011 Norwegian terror attacks in Oslo and on the island of Utøya, where 77 individuals lost their lives, including 69 youths on Utøya, a different trend emerges. While expressions of solidarity increased in France, they actually decreased over time in Norway (M. Eriksson, 2016). This discrepancy underscores the potential impact of the event's location on the observation of positive affective outcomes, such as solidarity expressions. Therefore, it is important for future studies to focus on events that are relevant in multiple locations to gain a more comprehensive understanding of the factors influencing social sharing and its emotional consequences.

In a similar vein, it is worth considering that studying acute events, where sharing predominantly occurs after the central event has concluded, such as the French and Norwegian terrorist attacks, as well as the Madrid bombings, may yield different outcomes compared to the examination of ongoing events characterized by chronic and enduring trauma. One plausible reason for expecting dissimilar results is that in ongoing crises, the information and intensity of the event can change dynamically, and individuals may have to adapt to new negative news and events (Reser & Swim, 2011; Yim, Lorenz, & Salkovskis, 2023). Conversely, in acute events, such as those that transpired within a single day (although their consequences naturally persisted for an extended period), there may be less requirement for individuals to adapt to evolving information. This might allow them to concentrate on regulating their emotions without the added burden of adjusting to new negative events.

There is one qualitative study that did investigate the consequences of social sharing during an ongoing negative event on social media, offering support for the presence of both

positive and negative outcomes, similar to what has been observed in acute events. Lazard et al. (2021) used semi-structured interviews to explore how cancer patients experienced both support and disheartenment after expressing themselves on social media. Although their study focused on individual experiences rather than collective ones, it sheds light on the complex nature of social sharing during times of adversity. It is unclear how this particular finding extends to collective events. In such collective events, emotions may be heightened compared to how individuals would feel about the event in isolation (Goldenberg, Garcia, Halperin, & J. J. Gross, 2020; Páez, Rimé, Basabe, Wlodarczyk, & Zumeta, 2015). Conversely, an alternative hypothesis suggests that when multiple individuals undergo the same event and perceive it collectively, emotions may be less intense as they have others to help regulate their feelings (Goldenberg, 2022a). As of now, there is a lack of definitive empirical evidence regarding the emotional consequences of social sharing in response to an ongoing collective crisis.

The research presented in this chapter focuses on social sharing related to the ongoing COVID-19 pandemic, a global event that has persisted for many months and continues to affect people's lives at the time of this discussion. This context is distinct from many previous negative events studied on social media in the context of affective content sharing and SSE because the pandemic is a chronic and global collective stressor. Therefore, it provides a unique opportunity to explore social sharing in contexts that have not been previously studied. Additionally, in the context of this thesis, this setting offers an opportunity to expand on previous research by exploring the creation and sharing of affective content in a negative, non-political context. As noted earlier, the thesis adopts a comprehensive definition of SSE, encompassing both the creation and sharing of online content related to negative events. The adoption of this inclusive approach was based on the

consistent findings between primary and secondary social sharing, with the exception of one non-significant effect that still trended in the same direction as the other findings. This divergence can be attributed to the smaller number of observations within and between participants in that particular subgroup. Consequently, in order to provide a concise summary of the findings, I decided to combine the results of primary and secondary social sharing. For a detailed breakdown of separate analyses, please refer to the Appendix of Chapter 4.

Research Questions and Goals

Social sharing on social media is extremely prevalent, especially in response to negative events, but there is a need for more empirical work to investigate whether sharing and creating crisis-related content truly serves the purpose of emotional recovery for the content producer. Furthermore, this aspect allows for an analysis of the prevalence of shared and created affective content *during* a negative event rather than *after* an event, as was investigated in Studies 2 and 3 of this thesis. In Study 4, I processed a large sample of tweets related to COVID-19 posted on Twitter to track how publicly communicated affect evolved over an extended period as users discussed the pandemic on social media. The data was collected from two Western countries, the United Kingdom (UK) and the United States (US), whose COVID-19 timelines had been similar at the start of the research (June 2020). This approach allowed me to assess how people were feeling approximated by their affective expressions before and during the outbreak of COVID-19 in a naturalistic setting, without the issues of recall bias that can characterize retrospective questionnaires. Twitter provided reliable longitudinal data for time-series analyses, unlike longitudinal survey studies which are more prone to participant attrition. This method was also appropriate for examining a

prevalent form of social sharing in the "new normal" where people came to spend much of their social life online due to social distancing rules in response to the pandemic.

To summarise, Study 4 had two main research objectives. First, to identify the prevalence of affective content that is created and shared during a sustained crisis. Second, to investigate whether discussing COVID-related events helped to regulate emotions or reactivate unpleasant feelings by analysing Twitter data produced approximately a month after COVID-19 was declared as a 'pandemic'. A panel of users' tweets that were related to the pandemic was collected and counted to provide a measure of pandemic-specific social sharing in general and for sub-periods within the month in question. Then, the sentiment analysis tool VADER (as also used in Study 1) was used to measure users' emotions expressed in their tweets (sentiment scores), which constituted my dependent variables. Sentiment scores were calculated for the whole one-month period and for sub-periods during that month to permit assessment of emotional effects over time.

Mixed effects of sharing have been reported in previous studies investigating acute collective crises (Garcia & Rimé, 2019; Rimé et al., 2010). This study assessed whether similar effects operate during a chronic crisis, or whether the balance between positive and negative effects of sharing differs because of the crisis's continuing nature. One possibility was that users who tweeted about the pandemic more often would tend to express more negativity and less positivity in their tweets than those who tweeted about it less often as a consequence of repeatedly revisiting COVID-related unpleasant feelings and thoughts. An alternative possibility was that users who tweeted more about COVID-19 would tend to express less negativity and more positivity as a result of the social support they solicited and the appreciation that they were not alone in facing this unprecedented situation. Finally, it is also plausible that users might exhibit a mixture of positive and negative consequences,

whereby both positivity and negativity increase simultaneously. In other words, my study addressed the question of whether higher levels of COVID-related sharing led to emotional worsening, improvement, or a combination of both.

Study 4: Affective Social Sharing During an Ongoing Crisis

Method

Participants. Following Garcia and Rimé's (2019) two-step procedure, I first sampled a panel of 32,592 users who used hashtags related to the COVID-19 pandemic in their original (i.e., not retweets), English tweets published between 20th April 2020 and 20th May 2020[4] via Twitter's application programming interface (API). Hashtags were selected by manual inspection[5]. The selected ones were commonly used to refer to the pandemic and avoid potentially ambiguous content. For example, #flu was also a common hashtag, but it had been used to refer to Avian influenza and seasonal influenza as well. As pointed out by Booten (2016), simply using hashtags to collect tweets on a certain topic for analysis is

[4] I initially chose 14th June 2020 - 17th July 2020 (the month in which we started conducting the present research) as the crisis period for investigation. However, I later came to realise that the death of George Floyd on 25th May 2020 might have confounded my data because it was a significant socio-cultural event that had further upset the world during the pandemic. Therefore, I altered my choice of the crisis period. 20th April 2020 - 20th May 2020 was a period after COVID-19 hit the UK and the US in January and triggered the first lockdowns in both countries in March. During this period, daily COVID case numbers were high but stable, and public emotions had not yet been affected by the Floyd tragedy.

[5] Hashtags that were selected after manual inspection: #Covid19, #Corona, #nCov2019, #Rona.

subject to sampling bias because the use of hashtags tends to change over time. Therefore, I used hashtags only for the purpose of initially identifying users who talked about COVID-19. I then downloaded *all* tweets published by each of these users (regardless of the hashtag) for analysis.

I also examined each user's Twitter profile and excluded those who left their location field empty or specified a location outside the UK or the US. I determined their location using their self-proclaimed location rather than their geo-tagged location since the latter was only available in 20% of all cases. With the aid of a manually curated dictionary containing UK and US cities, counties, and states, I identified these users as either UK or US residents. As many US cities share names with UK cities, I did not differentiate between these groups due to the possibility of overlapping. 'Verified' users (whose account is of public interest, such as brands and influential individuals) were excluded because I wanted to target ordinary users whose tweets are less likely to receive excessive public attention as shown in Chapter 2. Similarly, I excluded users with COVID-related terms as part of their usernames because they could potentially be accounts specifically intended to distribute COVID-related information. In addition, I was aware of the well-known presence of automated accounts ("bots") on Twitter. Previous studies have estimated that regular human users produce fewer than 100 tweets per month (see Table 1 in Haustein et al., 2016; Ottoni et al., 2014). I used the first 10% of my data to run a pre-registered pilot study to see how many tweets the current panel of users usually produced during the month. Since means are susceptible to outliers, I focused on the median number of tweets instead. Further details of this pilot study can be found in the Appendix of Chapter 4. Based on the results of this pilot study, I chose a pre-registered threshold for the exclusion of participants of 200 tweets (+1 *SD* from the mean) which allowed me to retain 85% of the sampled users. Any users passing the threshold were

excluded because a larger average number of tweets "increases the probability that the user is a bot" (Subrahmanian et al., 2016). The new median number of tweets after implementing this threshold was 50.

After revising my panel of users, I collected all of the tweets that they posted as well as shared and then filtered out users who did not publish any tweets during the crisis period. After applying these exclusion criteria, my final sample size was reduced to 29,484 users who published a total of 1,683,322 tweets during the crisis period. To identify tweets that focused on COVID-related events, I checked if the text contained any topic-related terms from a pre-defined dictionary[6,7]. This dictionary was developed by classifying the topics described in the open-ended responses provided by participants in Study 2's online questionnaires (reported below)[8]. As a result, 337,831 COVID-related tweets were included in my main analysis. All

[6] To validate the performance of my pre-defined dictionary at detecting COVID-related content in tweets, I randomly sampled 500 tweets from my pool of 1,683,322 tweets (regardless of their content) published during the crisis period and asked two independent judges to code these tweets for whether they perceived that a tweet talked about the pandemic in a post-hoc manner. Interrater reliability was r = .97. Results showed that the dictionary was able to correctly accept a COVID-related tweet when human judges perceived that it talked about the pandemic 87.61% of the time. It was also able to correctly reject a non-COVID-related tweet when human judges perceived that it did not talk about the pandemic 98.95% of the time. Further details are reported in the Appendix of Chapter 5.

[7] My pre-defined dictionary is described in more detail in the Appendix of Chapter 5. I also adopted a more extensive dictionary including more than 200 terms (Dimitrov et al., 2020) and repeated my VADER analysis with it. Results were consistent with those based on my pre-defined dictionary.

[8] A recent study by Okango and Mwambi (2022) has also used a dictionary-based method to screen tweets for sentiment analysis. Five out of the seven terms that they used also appeared in my dictionary, which contained ten terms in total.

procedures of this study were granted ethical approval by the Medical Sciences Division Ethics Committee (#R69862 and 69987) at the University of Oxford.

Measures. Users' level of social sharing and the sentiment of each of their tweets were the main dependent variables. I measured social sharing by counting the number of COVID-related tweets each user posted during the crisis period. For the sentiment of each tweet, I used VADER[9] to detect both valence (positive or negative) and emotional intensity (strength) of each word in the tweet. As mentioned previously, VADER considers the immediate linguistic context of the words used in a tweet and factors in the emphasis implied by capitalisation and punctuation. It generates a total sentiment score, a categorical variable (positive, neutral, negative) as well as positive, neutral, and negative sentiment scores by respectively summing up the intensity of each positive- and negative-affect word in the tweet. It's worth noting that the positive and negative dimensions in VADER are not entirely independent of each other, as discussed earlier in Chapter 2. In fact, separating these dimensions often leads to poorer model fit as in Studies 2 and 3 of Chapter 2. Yet, in contrast to all three previous studies, I used the separate VADER dimension for the analysis of the relationship between social sharing and positive and negative consequences despite this limitation for two reasons. First, it is conceivable that both positive and negative consequences can concurrently increase with heightened social sharing. If I were to average

[9] I repeated my analysis using the sentiment analysis tool used in Chapter 3, SentiStrength, and obtained consistent results with those found using VADER. Moreover, the latter was in fact confirmed to be a more suitable tool for sentiment analysis, as my model comparisons showed that it offered better model fit. Further details can be found in the Appendix IV.

the sentiment of each day using a continuous sentiment score, the positivity, and negativity could potentially offset each other. By employing a separate dimension, I can capture and retain the simultaneous increase in both positive and negative aspects. Second, I also compared VADER with SentiStrength (used in Studies 2 and 3) and found that VADER outperformed SentiStrength in the AIC model fit (see AIC across the reported results in Appendix of Chapter 4). Therefore, I present the VADER results in the main text of the thesis, while the SentiStrength results, which are largely consistent with those of VADER, can be found in the Appendix of Chapter 4. As in Study 1, I used the categorical VADER score to analyse the prevalence of affective content, while utilizing the continuous sentiment score to examine the relationship between sentiment and sharing behaviour.

Preregistration. The preregistered analysis plan for Study 4 is available at https://osf.io/dnsuj. This chapter does not present the results of all of the pre-registered analyses but concentrates on those that are most pertinent to evaluating the possible positive or negative consequences of social sharing as a whole[10]. The decision to combine the findings of primary social sharing (creating content) and secondary social sharing (sharing content) was made due to the similarity in results obtained from both types, as evidenced in the Appendix of Chapter 4. My analyses were designed to assess the hypothesis that sharing might have both positive and negative effects on sentiment, so I did not presuppose that sharing would be entirely beneficial. A non-preregistered Granger-Causality analysis was also conducted to clarify the implications of the main pre-registered analysis. In addition to

[10] The pre-registered analyses using LIWC (Pennebaker et al., 2001) to estimate emotionality of tweets are reported in the Appendix of Chapter 4. Results are not substantially different from the results using VADER reported in the *Results* section.

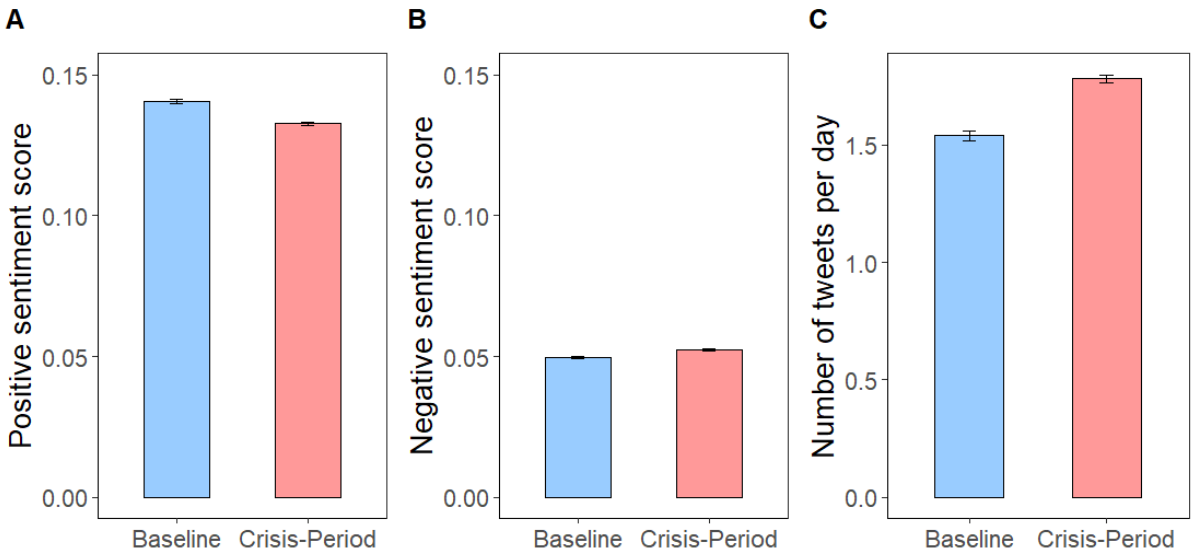
the results presented in the manuscript and registered in the pre-registration, I also analysed the prevalence of created and shared affective content in order to expand upon the findings presented in Chapters 2-3. All data and analysis codes relating to the presented findings are available on the Open Science Framework (OSF) at https://osf.io/hwj63/.

Results

Influence of the Crisis on Affective Content Creation. I first evaluated the assumption that the general affective expressions online were negatively impacted by the pandemic during my chosen crisis period (20th April 2020 – 20th May 2020). For each user, I averaged sentiment scores of both positive and negative sentiment scores of all kinds of tweets (COVID-related and not) published during the crisis period and the baseline period (1st – 31st October 2019) for comparison.

A paired t-test showed that positive sentiment scores were significantly lower (t (29103) = -20.29, $p < .001$; see Figure 4.1A) and negative sentiment scores were significantly higher (t (29103) = 11.96, $p < .001$; see Figure 4.1B) during the crisis period (M_{pos} = 0.13, SD_{pos} = 0.053; M_{neg} = 0.049, SD_{neg} = 0.033) than during the baseline period (M_{pos} = 0.14, SD_{pos} = 0.067; M_{neg} = 0.052, SD_{neg} = 0.042), consistent with my expectations. In addition, I found that there was a significant rise in the number of tweets per day during the crisis period (t (29103) = 23.49, $p < .001$; see Figure 4.1C), which indicated that the pandemic (M = 1.78, SD = 1.82) was associated with greater daily activity by Twitter users compared to the baseline period (M = 1.53, SD = 2.03). To summarise, users tweeted more and expressed less positive sentiment and more negative sentiment during the crisis period than before it.

Figure 4.1 *Sentiment Scores and Number of Tweets Posted During Baseline and Crisis Periods*



Note. *Panels A and B respectively compare the average positive and negative sentiment scores of all kinds of tweets between the baseline (the left bar in every panel) and the crisis (the right bar) period. Panel C compares the number of all kinds of tweets per day between the two periods.*

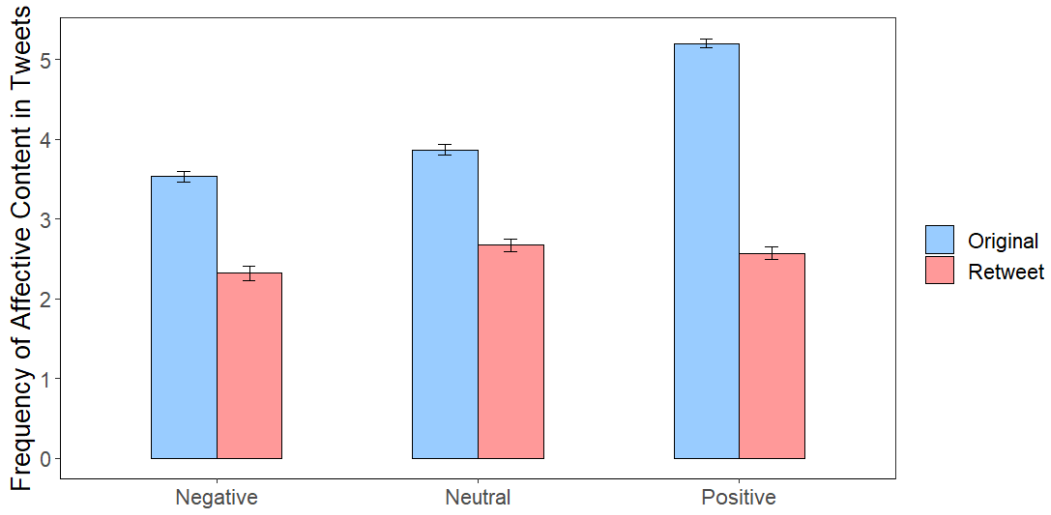
Created and Shared Affective Content during an Ongoing Crisis. Next, I investigated the content that users generated and shared about the crisis during the crisis period. To do so, I used the VADER categorisation to predict the average number of created and shared tweets of a user. Both the number of original tweets and retweets were left-skewed. In prior analyses involving count data such as the number of retweets in Studies 1-3, a log-modulus transformation was employed to address the presence of extreme values. Contrary to these previous analyses, I opted for an alternative approach. Rather than transforming the distribution, I applied a Poisson regression on the untransformed data. This was possible since the dataset, after excluding all users without tweets, did not contain zero

values. One significant advantage of this approach is interpretability. The absence of a transformation allows for the raw data to be read directly, thus eliminating the need for reverse transformation to understand real-world numbers. This approach provides a more intuitive understanding of the count data, and it simplifies the interpretation of the regression coefficients. Therefore, a Poisson distribution could accurately model the data while avoiding the computational issues associated with zero counts. To compare tweet frequencies within and between categories, I added the two tweet types (original and retweets) as an interaction factor to my analysis. As users with more followers tend to generate less negative content as demonstrated in Study 1, I also added the number of followers as a covariate.

In line with my findings from Chapter 2, I found that even in an ongoing crisis negative affective content was generated less frequently (M = 3.49, SD = 4.20) than positive affective content (M = 5.16, SD = 6.30, b = 1.67 [1.58, 1.75], SE = 0.045, t (92071.00) = 36.86, p < .001, R^2 = 0.073, as see Figure 4.1), and neutral affective content (M = 3.84, SD = 4.94, b = 0.33 [0.24, 0.43], SE = 0.047, t (92071.00) = 7.13, p < .001, R^2 = 0.073). Interestingly, users also shared less negative affective content (M = 2.34, SD = 2.30) than positive affective content (M = 2.65, SD = 2.68, b = 0.25 [0.13, 0.37], SE = 0.062, t (92071.00) = 4.03, p < .001, R^2 = .073), and the same was true for neutral affective content (M = 2.72, SD = 2.81, b = 0.35 [0.22, 0.47], SE = 0.062, t (92071.00) = 5.61, p < .001, R^2 = .073). When analysing the interaction of tweet type and affective content category, I found that positive affective content was relatively less frequent in retweets compared to originals (b = -1.41 [-1.56, -1.26], SE = 0.077, t (92071.00) = -18.35, p < .001, R^2 = .073), while there was no significant difference in the frequency of negative affective content compared to neutral content between originals and retweets (b = 0.013 [-0.14, 0.16], SE = 0.078, t (92071.00) = 0.17, p = .86, R^2 = .073). Overall users tended to create (M = 4.25, SD = 5.38)

more tweets than they shared about COVID-19 (M = 2.59, SD = 2.63, b = 1.75 [1.69, 1.81], SE = 0.031, t (92071.00) = 56.01, p < .001, R^2 = .073).

Figure 4.2 *Frequency of Affective Content Types for Original and Retweeted Content*



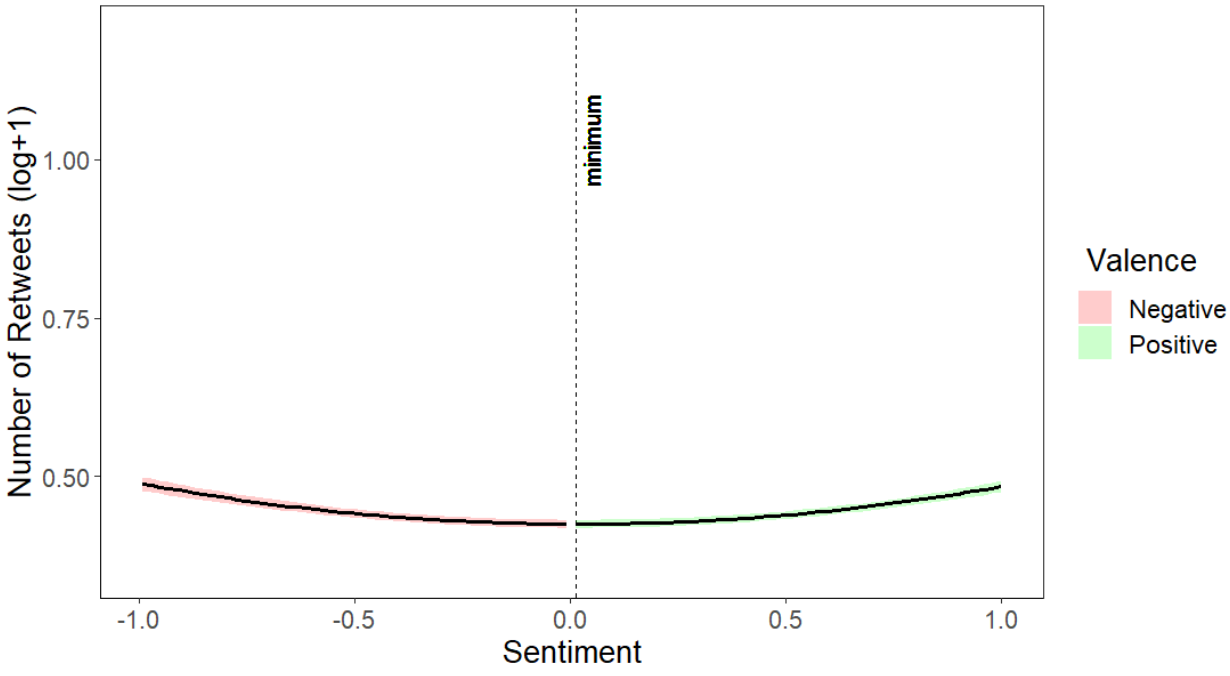
Note. *In both tweet types (originals, blue, and retweets, red) the least frequent affective content type was negative. Users created more tweets regarding COVID-19 than they retweeted.*

Associations between Sentiment Scores and Number of Retweets of Users' Original Tweets. In order to confirm the findings of Chapter 2 regarding the sharing patterns of original negative and positive affective content among ordinary users, excluding political events (as observed in Chapter 3), I conducted a replication analysis using the exact same methodology as in Chapter 2. My dependent variable was content sharing, measured by the log+1 transformed the number of retweets. As the predictor, I again used the continuous sentiment score from VADER (-1 extremely negative to +1 extremely positive) fitted using a quadratic mixed model to predict the number of retweets. As a reminder of the explanation previously given in Chapter 2, it is worth noting that in the quadratic function ($y = ax^2 + bx +$

c), a larger *a* coefficient in the present model indicates a stronger association between the strength of both positive and negative sentiment and content sharing. On the other hand, a larger *b* coefficient would suggest that positivity was shared more than negativity.

Higher sentiment scores were positively associated with the number of retweets for original content produced (*a* = 0.18 [0.17, 0.19], *SE* = 0.0067, *t* (845141.16) = 27.17, *p* < .001, R^2 = .085, see Figure 4.3). An increase in negativity led to a similar increase in the number of followers than an increase in positivity (*b* = -0.0028 [-0.0076, 0.0020], *SE* = 0.0024, *t* (845324.79) = -1.15, *p* = .25, R^2 = .15).

Figure 4.3. *Number of retweets as a function of sentiment for original content during the crisis*

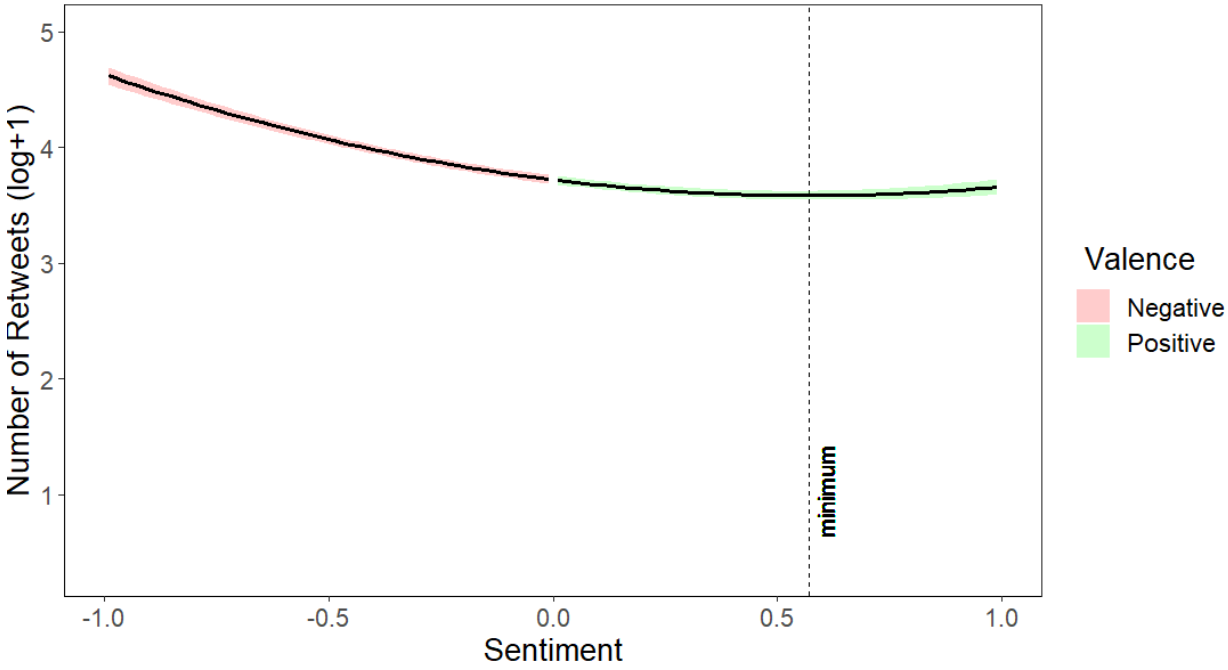


Notes. *The results suggest that stronger sentiment scores are associated with more retweets during the crisis period. The local minimum is reached at approximately the 0 sentiment score (neutral content), indicating that negative and positive affect are equally increasing the likelihood of retweets.*

Associations between Sentiment Scores and Number of Retweets in Users' Retweeted Tweets. As I collected all the tweets a user shared and created during the crisis period, instead of only the original as in Study 1 – 3, I can now also analyse the properties of the content that other users retweet. Specifically, my investigation focused on the relationship between the sentiment score of a retweet and its overall level of sharing. However, analysing retweeted content introduces several unknown factors, including the number of previous shares, the number of followers of the previous sharer or creator, and the nature of the relationship between the two. Despite these challenges, it seemed worth conducting this analysis in order to gain a comprehensive understanding of the content shared and created online. To analyse this question, I used the exact same analysis as in the previous paragraph using the sub-sample of retweeted content.

In retweeted content, sentiment scores were again positively associated with the overall number of retweets (a = -0.48 [-0.51, -0.45], SE = 0.017, t (84187) = -27.68, p < .001, R^2 = .087, see Figure 4.4). In contrast to the original content, the negative affective content had more retweets compared to positive affective content during the COVID-19 pandemic (b = 0.42 [0.35, 0.49], SE = 0.034, t (84187.00) = 12.29, p < .001, R^2 = .15).

Figure 4.4. *Number of retweets as a function of sentiment for retweeted content during the crisis*



Notes. *The results suggest that stronger sentiment scores are associated with more retweets for ordinary users during the crisis period. The local minimum is reached with a more positive affective tweet, indicating that negativity is more strongly positively associated with the number of retweets. In other words, negativity increases sharing more than positivity does.*

Linear Growth Models of Sentiment Change Over Time. After analysing the prevalence of affective content, I now turn to the analysis of the consequences of SSE for sharing users. I ran linear growth models to test the association between users' levels of pandemic-specific sharing and the sentiment scores of their COVID-related tweets over the sampled days of the crisis period. In accordance with my pre-registration document, the number of COVID-related tweets both created and shared, which was used to measure social sharing level, was log + 1 transformed, since the distribution was positively skewed. The

models were multileveled because positive and negative sentiment scores were nested within days and participants. Given that the act of writing and sending tweets itself may have an impact on affect regardless of their specific content, I included the *overall* number of tweets (COVID-related and non-COVID-related) posted by each user during the crisis period as a covariate in the models. The other covariate was *baseline*[11] sentiment scores (positive or negative sentiment depending on the valence in question) of tweets posted prior to the crisis period. The resulting model equation is as follows: *Sentiment Score ~ Social Sharing Level * Time + Baseline Sentiment + Overall Number of Tweets + (1 | User).*

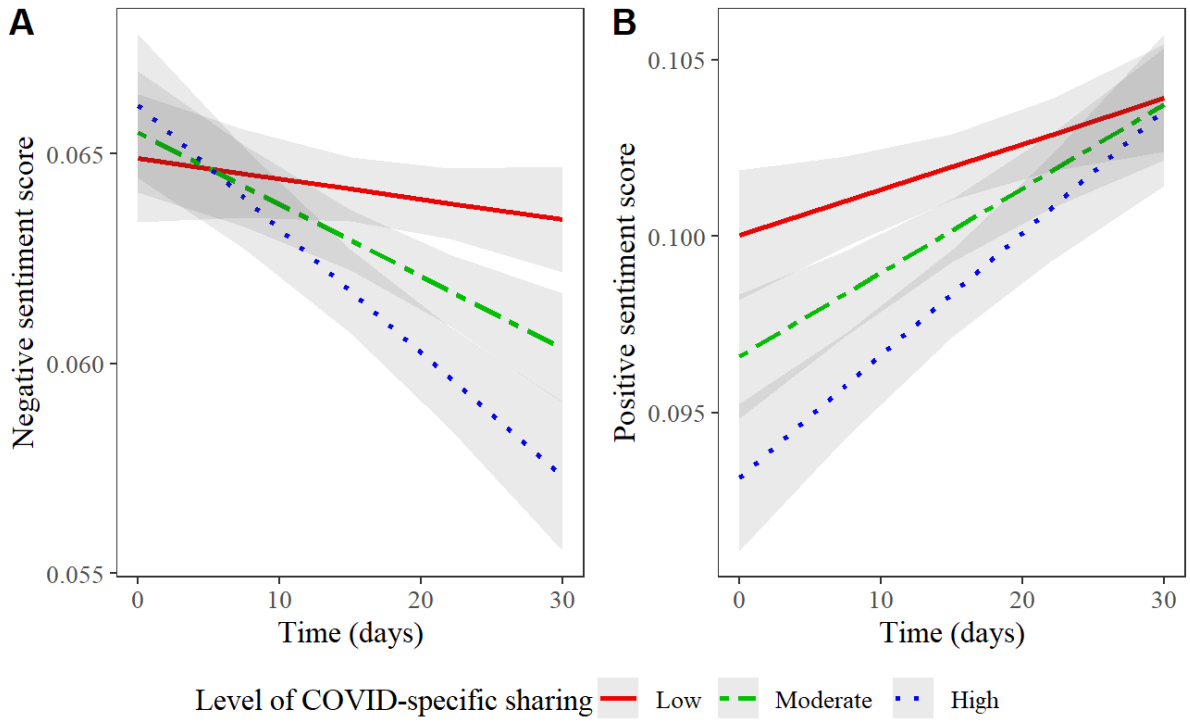
Negative sentiment scores were significantly positively predicted by the covariates representing baseline negative sentiment (b = 0.0156 [0.0151, 0.0160], SE = 0.00024, t (25219.036) = 65.67, $p < .001$) and the overall number of tweets (including COVID-related and non-COVID-related tweets: b = 0.0035 [0.0029, 0.0041], SE = 0.00031, t (68411.07) = 10.97, $p < .001$), showing that users who usually expressed more negative sentiment and who were generally more active on Twitter tended to have more negative things to say about the pandemic specifically. Time also had a significant negative main effect (b = -0.0015 [-0.0022, -0.0008], SE = 0.00035, t (52.59) = -4.28, $p < .001$), showing that the negative sentiment expressed in COVID-related tweets generally tended to decrease over the course of the study period. The final significant main effect was a negative effect of pandemic-specific sharing frequency (b = -0.0014 [-0.0019, -0.0009], SE = 0.00026, t (32699.57) = -5.17, $p < .001$), showing that after controlling for the covariates tweeting about the pandemic was generally associated with lower negative sentiment. Importantly, there was also a significant

[11] I collected the same panel of users' tweets published from 1st to 31st October 2019 (1,740,368 tweets) to provide baseline measures. For additional details of this baseline period, please see the Appendix of Chapter 4.

interaction between COVID-specific tweets and time (b = -0.0012 [-0.0014, -0.00078], SE = 0.00019, t (273529.81) = -6.37, p < .001), showing that the level of negativity in COVID-related tweets decreased over time to a greater extent for users who tweeted specifically about the pandemic more frequently than for users who tweeted specifically about the pandemic less frequently (see Figure 4.5).

Positive sentiment scores mainly showed the opposite pattern of effects. First, positive sentiment scores were significantly positively predicted by the covariate representing baseline positive sentiment (b = 0.0151 [0.0146, -0.0023], SE = 0.00028, t (25931.17) = 52.94, p < .001) but significantly negatively predicted by the overall number of tweets (b = -0.0030 [-0.0037, -0.0023], SE = 0.00038, t (73056.88) = -7.86, p < .001), showing that those participants who usually expressed greater positive sentiment and who tweeted less during the study period expressed greater positive sentiment. Second, there were again significant main effects of time (b = 0.0021 [0.0013, 0.0029], SE = 0.00042, t (62.74) = 4.90, p < .001) and of pandemic-specific sharing frequency (b = -0.0017 [-0.0023, -0.0010], SE = 0.00032, t (32786.73) = -5.01, p < .001), showing that positive affect generally increased over time during the study period and was lower for those who tweeted more about COVID. Finally, there was again a significant interaction between pandemic-specific sharing frequency and time (b = 0.0011 [0.00049, 0.0013], SE = 0.00020, t (275182.81) = 4.47, p < .001), showing that the level of positivity in COVID-related tweets increased over time to a greater extent for users who tweeted specifically about the pandemic more frequently than for users who tweeted specifically about the pandemic less frequently (see Figure 4.5).

Figure 4.5. *Sentiment Scores of COVID-Related Tweets as a Function of Time and Users' Level of Pandemic-Specific Sharing*



Note. *Panel A shows negative sentiment scores as a function of time and users' level of pandemic-specific sharing while Panel B shows positive sentiment scores. The level of pandemic-specific sharing was a continuous variable but is presented as three categories in this figure to allow clearer visualization of the interaction effect. Different levels of sharing respectively modelled users who tweeted M – 1 SD times (minimum), the mean number of times (moderate), and M + 1 SD times (maximum).*

Granger-Causality Tests. Having established that social sharing across the full study period was associated with decreasing negativity and increasing positivity as the study month progressed, I next focused more directly on the temporal effects of prior social sharing levels on subsequent emotionality of tweets in sub-periods of the month. I performed Granger-Causality tests to answer the question of whether social sharing was an antecedent predictor of the emotionality of tweets (Granger, 1969). According to Granger (1969), temporal

causality is confirmed if one-time series in a more complex model is better at predicting a later time series than in a simple model.

In the context of the present study, I wanted to know whether sentiment scores of COVID-related tweets posted each week[12] (*T*) could be predicted by the previous week's (*T* - 1) pandemic-specific sharing frequency. Therefore, in the simple model, I entered sentiment scores of COVID-related tweeted at time point *T* - 1 as the *basic* predictor (the variable on which an outcome variable necessarily depends), working on the assumption that how users felt in a given period has continuity with how they felt in immediately preceding periods. For the complex model, I then added pandemic-specific sharing frequency at time point *T* - 1 as the second predictor. This was the *key* predictor to be tested for its ability to forecast the outcome variable. The equations representing the simple and complex models were thus as follows:

***Simple model*:** *Sentiment of COVID-Related Tweets on T ~ Sentiment of COVID-Related Tweets on T - 1 + Baseline Sentiment + Overall Number of Tweets + (1 | User)*

***Complex model*:** *Sentiment of COVID-Related Tweets on T ~ Sentiment of COVID-Related Tweets on T - 1 + Number of COVID-Related Tweets on T - 1 + Baseline Sentiment + Overall Number of Tweets + (1 | User)*

[12] I also repeated the Granger causality test after resetting the time period to 2 weeks (with the time lag remaining as 1), meaning that I aggregated my data in groups of every two weeks and aimed to predict sentiment scores of next two weeks' tweets using the last two weeks' pandemic-specific sharing frequency. Given that there were only four weeks throughout the crisis period, what I did was simply split the data in half and use the first two weeks of tweets to forecast the second two weeks of tweets. Results were consistent with when the time period was set to 1 week. Further details are reported in Appendix of Chapter 4.

The null hypothesis of the Granger-Causality test is that the simple and complex models are not significantly different, or in other words that the additional predictor included in the complex model is not useful for forecasting the outcome variable. Therefore, if the complex model is shown to have a significantly better fit than the simple model, then I can say that the emotionality of tweets at time point *T* is 'Granger-caused' by social sharing at time point *T* - 1.

It is worth noting that the study only covered four weeks during the crisis period, and not all users tweeted about the pandemic across all four weeks. Since I was interested in the immediate effects of the previous week on the next for each user, I excluded the weeks in which the user engaged in no social sharing and the weeks immediately after those weeks from my analysis. For example, if a user did not tweet about the pandemic on the third week, I removed their data from that week and the fourth week.

Moreover, because sentiment scores were nested within participants, in both the simple and complex models I included a random intercept for each user, as well as the overall number of tweets (COVID-related and not) posted by each user during the crisis period and baseline emotionality as covariates (as I did for the linear growth models reported above). A traditional Granger-Causality test only includes two predictors, so by adjusting my models in this manner, I was, in fact, conducting a *Granger-Causality-inspired multileveled test* (see for a full example Green-Pedersen, Mortensen, & Thesen, 2017). However, since the same adjustments were applied to both models, all aspects of the two models except for the additional predictor in the complex model remained constant, allowing me to isolate the impact of the key predictor.

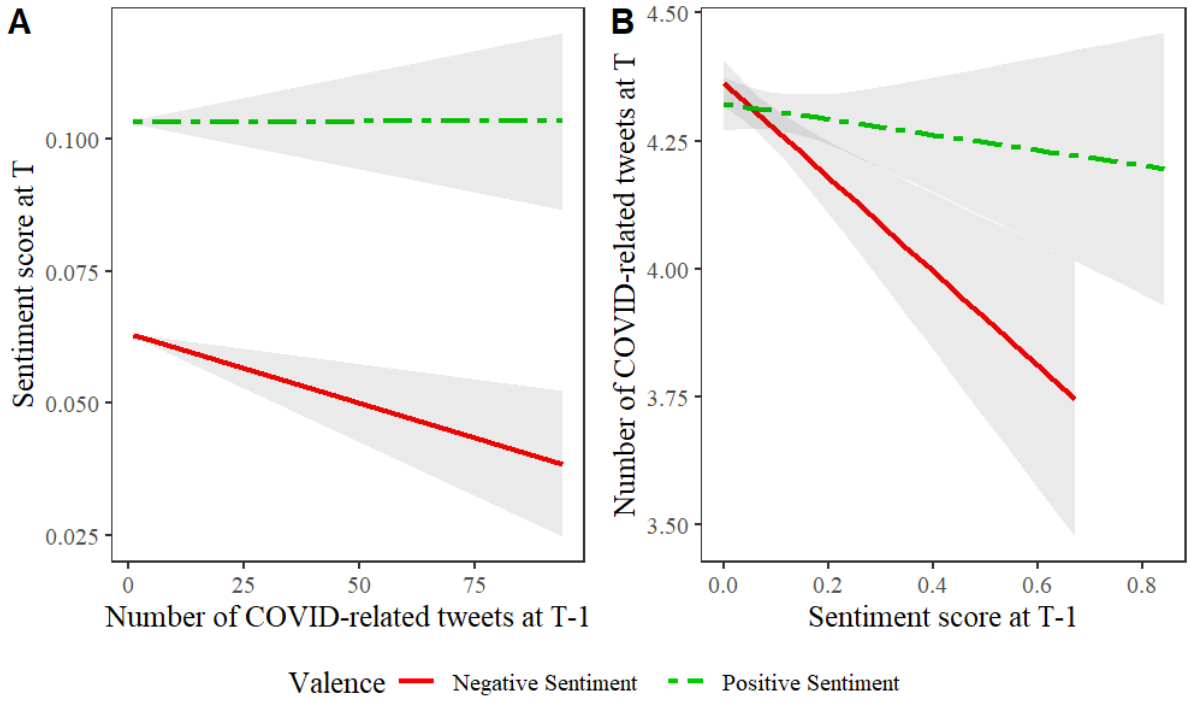
I compared my two models for positive and negative sentiment scores separately. Results showed that the complex model had a better fit than the simple model ($\chi^2 = 12.99$, $p <$

.001) in the case of negative sentiment scores, indicating that a higher level of pandemic-specific sharing in the previous week 'Granger-caused' a decrease in the use of negative-affect words in COVID-related tweets in the following week (*b* = -0.00028 [-0.00041, -0.00011], *SE* = 0.000077, *t (*29901.48) = -3.63, *p* < .001; see Figure 4.6A). Although the null hypothesis of the test was rejected, the possibility of reverse causation remained. Thus, I ran the test again in reverse, to check if negative sentiment scores of COVID-related tweets at time point *T* - 1 led to any change in pandemic-specific sharing frequency at time point *T*. Temporal causality was indeed detected in this reverse direction[13]. I, therefore, compared the fits of both directional models and found that the forward-causation model (*AIC* = -159187) had a better fit than the reverse-causation model (*AIC* = 309346). Therefore, I deemed that the influence of prior social sharing on subsequent negativity was greater than that of prior negativity on subsequent social sharing.

In the case of positive sentiment scores, the Granger-Causality test did not yield any significant results for forward causation (*p* = .98). Reverse causation was also non-significant (*p* = .40).

[13] The simple model (with pandemic-specific sharing frequency at time point *T* - 1 as the basic predictor) was shown to be significantly different from the complex model (with negative sentiment scores of COVID-related tweets at time point *T* - 1 as the key predictor). The latter had a better fit than the former (χ^2 = 16.86, *p* < .001), indicating that higher negativity in the previous week 'Granger-caused' a lower level of pandemic-specific sharing in the following week (*b* = -0.061, *SE* = 0.014, *t*(57072.06) = -4.10, *p* < .001, 95% CI [-0.097, -0.028]; see Figure 4.6B).

Figure 4.6. *Temporal Causality Between Sentiment Scores and Social Sharing of COVID-Related Content*



Note. *Panel A illustrates how pandemic-specific sharing frequency in the previous week predicted sentiment scores of COVID-related tweets posted during the following week (i.e., forward causation). Panel B, on the other hand, illustrates how sentiment scores of COVID-related tweets posted during the previous week predicted pandemic-specific sharing frequency in the following week (i.e., reverse causation). T - 1 stands for a time period (1 week) going back by a time lag of 1—that is, the week before. The darker lines represent positive sentiment scores of COVID-related tweets and the lighter lines have negative sentiment scores.*

Discussion

In Study 4, I answered two types of questions. First, what valence of content is shared and created by ordinary users during an ongoing negative event, and second, what is the consequence of sharing and creating online affective content for users? Beginning with the

prevalence of affective content, I conducted analyses investigating the content that users created and shared as well as how emotional intensity influenced how their own content was shared. I found that users created more content during the pandemic than before it and that both original and retweeted content were predominantly positive. Although this finding replicates Study 1 by showing that original content is more positive than negative, it is still surprising that users maintain their emphasis on positive sentiment during negative events. In line with the assumption that the pandemic has a negative effect on people's expressions, I found that positive sentiment decreased while negative sentiment increased. This still contrasts with the findings of Studies 2 and 3, where negative affective content was both created and shared more frequently during negative events.

When analysing how the intensity of the original COVID-19 content posted by users affected the likelihood of others sharing their content during the crisis, I discovered that both positive and negative affective content were equally likely to be shared by others, replicating the finding of Study 1 with regard to ordinary users. This again is in contrast to the findings of Studies 2 and 3, where an increase in negativity was found to be more strongly associated with an increase in retweets than positivity in the context of politics. These results imply that, for ordinary users outside of politics, there is remarkable positivity even during ongoing negative events. This evidence suggests that social media may not be as inherently negative as initially assumed and detailed in Chapter 1. However, upon investigating the total number of retweets posted by ordinary users, I did find that negativity tended to be associated with more sharing. This implies the presence of mechanisms beyond the commonly proposed negativity bias, which posits that negativity is shared primarily due to increased attention. If the negativity bias were solely responsible, it would necessitate the assumption that users not only share more negative content but also have their own negative affective content shared

more frequently by others. I did not find sufficient evidence to support this assumption. In conclusion, with regard to the prevalence of affective content, it can be inferred that while negativity increased and positivity decreased during the pandemic, there is no evidence to suggest that negative affective content is more viral or prevalent than positive affective content for ordinary users.

My investigations into the consequences of SSE during the ongoing crisis showed that tweeting about pandemic-related content more frequently was associated with decreasing use of negative-affect words and increasing use of positive-affect words in pandemic-related tweets over the time period of the study. In other words, sharing more pandemic-related content seemed to help Twitter users feel less negative and more positive about the pandemic, consistent with an emotional-recovery hypothesis.

The results of the Granger-Causality tests are consistent with the conclusion that there was temporal causality between social sharing during the pandemic and users' expression of negative sentiment. In particular, the number of COVID-related tweets in the previous week negatively predicted the negativity of users' COVID-related tweets in the following week across the four weeks of the study. This finding helps to rule out some of the alternative explanations for the over-time effect of COVID-specific tweeting on expressed negative sentiment. In particular, the better model fit for the lagged effect of $T - 1$ sharing on sentiment scores at T than for the lagged effect of $T - 1$ sentiment scores on sharing at T makes it less likely that the sharing by time interaction in my main analysis depended on reverse causation (i.e., users tweeted less frequently as a consequence of greater prior negative affect). Further, any general effect of dispositional characteristics associated with the propensity to engage in COVID-related tweeting cannot easily explain the pattern of temporal associations. For example, it seemed possible that tweeting more about the pandemic reflected greater

sensitivity to concerns relating to COVID or illness more generally, and that this sensitivity rather than tweeting itself explained some of my reported effects. However, such an association cannot easily explain the lagged effects which are based on changes in the frequency of tweeting rather than overall levels of tweeting. Of course, it remains possible that changes in COVID-related tweeting reflected changing levels of COVID-19 sensitivity, and this third variable also contributed to the reported temporal effect. Ruling out all possible third-variable explanations clearly presents practical issues for this kind of online research, but I still believe that the most likely causal explanation for my findings is that sharing of COVID-related content improved affect over time.

Another interpretational issue arises from the fact that initial levels of positive sentiment were lower for users who shared more COVID-related content (see Figure 4.5). Thus, the improvement in their positive sentiment over time may simply have reflected the fact that they started with atypically low levels and returned to a baseline level over time because of regression to the mean (e.g., Barnett, Van Der Pols, & Dobson, 2005). Further, hedonic adaptation (e.g., Frederick & Loewenstein, 1999; Wilson & Gilbert, 2008) or regulatory equilibrating processes (e.g., Cummins & Nistico, 2002; Headey & Wearing, 1989) may have operated to restore baseline levels of positive affect regardless of the extent of sharing. In either case, the reported change in expressed positive sentiment might have reflected differences in this variable that were confounded with initial sharing levels rather than the effects of sharing itself. Because the Granger-Causality test also failed to support any significant lagged effect of COVID-specific sharing on subsequent positive sentiment, I cannot rule out the possibility that the reported temporal effects on positive sentiment depended at least partly on the different levels of positive sentiment that participants experienced at the beginning of the sampled time period.

By contrast, the significant main effect of sharing showed that negative sentiment scores tended to be lower across the study period for those who shared more content about the pandemic than for those who shared less (except during the first few days when they were approximately equal, see Figure 4.5), and this difference increased in size as time progressed as evidenced by the significant time by sharing interaction effect. Further, the Granger-Causality test showed that COVID-specific sharing in each week of the study was a significant negative predictor of negative sentiment in the following week. This means that negative sentiment continued to decrease even after higher-sharers were tweeting less negative content than lower-sharers (from week 2 onwards). This pattern of results is difficult to explain in terms of regression to the mean, hedonic adaptation, or affective equilibration. Instead, the greater reduction in expressed negative sentiment over time for users who shared more COVID-related content is more likely to reflect the alleviating effects of sharing on negative affect.

In contrast to the previous chapters, I treated sentiment scores as an indirect proxy for affect. As Steinert (2021, p. 60) noted, "expressing emotions online does not mean that these emotions are inauthentic or do not represent what people really feel." In other words, Twitter provided me with data on what users wrote when they were expressing their feelings about the crisis, and VADER helped me estimate their affective states at that moment by generating sentiment scores based on the emotionality of the words used in their texts. Given that sentiment scores provide only an indirect measure of users' experienced emotions, Gloria Wai Ma and I included a second study in this paper's manuscript to supplement the online findings. This study aimed to directly ask participants to rate their affective states in questionnaires. The data from both studies, which used different methods, suggested that

sharing COVID-related negative experiences on social media led to an improvement in participants' affective states rather than worsening them.

While Twitter data provided valuable insights into social sharing during the COVID-19 pandemic, there were two pieces of fine-grained information that it could not provide. First, it was unclear whether the emotional recovery goal was partly achieved after initial social sharing or depended on a longer series of discussions. Second, it was uncertain whether similar outcomes would be observed in offline communications, such as in-person conversations. Therefore, we also compared face-to-face (offline) and remote (online) social sharing to provide a more comprehensive understanding of the phenomenon. These aspects were addressed in another study conducted by fellow DPhil student, Gloria Wei Ma. Despite its smaller sample size, this study revealed selective effects of sharing on negative emotions related to personal experiences, showing that participants tended to feel less negative about such events only when they had talked about them. Overall, across studies, the nature of emotional improvement was similar: social sharing appeared to help people feel less negative about what happened.

The findings in regard to the first question of the thesis about the prevalence of affective content demonstrate that the negativity of an event is not sufficient for negative content to become viral among ordinary users, who surprisingly displayed resilience and positivity during difficult times. Turning to the second research question—specifically, the consequences of affective content—the analysis, in conjunction with an additional survey study, contribute to the literature on social sharing and emotion regulation by highlighting the benefits of reaching out and conversing with others about negative emotions during times of distress. Additionally, this research serves as an example of how combining different

methodologies can provide a fuller understanding of the effects of sharing by compensating for their respective limitations.

After providing evidence for the potential benefits of expressing and sharing affective content, an important question that remains unanswered is the potential consequences of seeing such tweets. These consequences not only include the well-being of the perceiver but also how they perceive the emotional intensity of events and news, such as updates on the COVID-19 pandemic, after being exposed to multiple users discussing the topic. The next chapter of this work aims to address this question by investigating how individuals make judgments about emotional sequences and how this may ultimately influence their own well-being.

CHAPTER 5. THE IMPACT OF NEWS FEEDS ON THE PERCEPTION OF AFFECTIVE NEWS ARTICLES AND OTHERS' RESPONSES

Abstract

The presentation of online content typically follows a sequential format, akin to the *news feed* or timeline format used by social media platforms. As highlighted in the previous chapters this content often uses emotional language. In this chapter, I explore how individuals perceive affective content when consuming it in a sequential manner and examine how this perception influences their intentions to share such content. Through a series of four experimental studies, I illustrate that people tend to overestimate the intensity of emotional language found in consecutively presented responses to fictional news articles. Additionally, I demonstrate that this estimation is influenced by the context of the sequence, in this case, fictional news articles, and that the estimation of news article's intensity can, in turn, be influenced by the estimation of the response sequences. The news articles that are perceived as being more emotional, are more likely to be shared; however, there is no difference in sharing between positive and negative affective content. Finally, I delve into the mechanism behind this overestimation. I suggest that the reasons that people overestimate may not be that they have preferential memory for more intense responses, but that they perceive these responses as more representative or diagnostic than neutral content. The overall objectives of this chapter are twofold. First, I establish that sequential presentation leads to an overestimation of both the contextual intensity and people's responses to it. Second, I validate the assertions made in Chapters 1-4 by demonstrating that stronger affective content is more prone to being shared.

Introduction

Every day, the average American person spends over two hours browsing online (Perrin, 2015), scrolling through 300 ft (91.44 m) of messages per day (Wade, 2017). They peruse this stream of content mainly to seek information and stay up-to-date on public sentiment regarding topics such as politics, health, news, and science, in order to shape their own opinions (Kelly & Sharot, 2023; Shearer & Grieco, 2019). On social media platforms, for example, users often come across and engage with affective content including news articles, along with the accompanying responses from other individuals. In earlier chapters, I have presented evidence demonstrating how frequently users express and share a significant amount of affective content. To understand the overall sentiment of the articles and people's opinions, users must not only decode emotions from multiple texts, but also aggregate them into a summary representation such as an average. Yet, despite the ubiquitous and frequent engagement with content presented in a feed-like format, which exposes users to sequences of affective content, there has been limited empirical research investigating the perception of affective content in such formats or the consequences of viewing such content.

The research presented in this chapter aims to address two significant gaps in the literature. First, I investigate how feed-like patterns of information shape our perception of news and others' emotions towards it. Second, I aim to explore the underlying perceptual processes involved in aggregating affective content. My experiments specifically focus on the perception of affect in texts, given that emotional language is a central driver of engagement with content on social media (Goldenberg & J. J. Gross, 2020; Schöne et al., 2022; Schöne et al., 2021), which I have also demonstrated in previous chapters. Individuals encounter an average of 64,986 words per day online (Brinberg et al., 2021), and I argue that considering this volume of content and users' limited attention capacity, it is likely that they tend to pay

more attention to affective content (Zhang & Qu, 2018). As a result, when affective sequences are aggregated, perceivers may overweight stronger expressions, leading to an overestimation of the sequence's average emotional language intensity.

Affective Content on Digital Media

Digital media is rich with affective content which seems to impact both user engagement and well-being (Kramer et al., 2014; Schöne et al., 2022). Unlike traditional media, digital media typically allows users to gather information from people's responses and reactions to a specific piece of news. This raises the question: How do users estimate the average emotional expressions of those responding sequentially to news items? Further, to what extent does this estimate influence users' own valuation? Existing research indicates that social media has the potential to amplify our perception of events, opinions, and emotions (for a review, see Tucker et al., 2018), but it remains uncertain whether the way we consume content and aggregate information on social media contributes to the misinterpretation events, opinions, and emotions. Considering that social media has approximately 2 billion active users (Statista, 2022a, 2022b), it is important to understand this aggregation process as well as its consequences.

Previous research suggests that people may overestimate the intensity of emotional language in online texts compared to the content producers' intention. Brady et al. (2022) conducted a study in which they asked participants to rate the intensity of outrage and happiness in tweets, and compared these ratings to the authors' intended intensity ratings. The study found that readers, especially heavy social media users, consistently overestimated the intensity of affective content compared to the authors' self-rated emotions. They also found that people overestimated the total intensity of sequences of content that expressed

outrage, but only in the condition where there was a large discrepancy between authors' intended emotions and users' perceived emotions. In this study, the overestimation of the sequence was calculated using the average ratings of each text as estimated by previous raters as the objective mean, subtracted by the participants' ratings of the sequence. However, it is uncertain whether this overestimation was a result of the way they calculated their objective average or the process of aggregating these ratings by the reader. Since the participants who read the sequence differ from the individual text raters, the estimations could vary due to the fact that previous raters have seen a greater number of items in total to compare with. In other words, the calculation of the objective mean may not accurately reflect how the sequence participants *would have* rated the individual texts. For example, even in a hypothetical scenario where participants are able to accurately incorporate sequential information into a mean, using independent raters to calculate the objective mean would still yield an overestimation, if the previous raters perceived the texts as more intense than the participants who observed the sequence.

By contrast, some studies suggest that people may underestimate the emotional language intensity of sequences of texts due to affective habituation, which is when individuals fell less intense in response to emotional stimuli over time (O'Brien & Kassirer, 2019). As users scroll through their social media feeds and come across affective posts, they may become desensitized to their affective content on repeated exposure (Campbell, O'Brien, Van Boven, Schwarz, & Ubel, 2014). This effect has been observed on Twitter in response to content concerning COVID-19 (Stevens, Oh, & Taylor, 2021), and gun violence (Jianing Li, Conathan, & Hughes, 2017), with users showing decreased affective responses to these negative events. However, it is unclear whether affective habituation influences how

individuals evaluate sequences of affective texts, as the concept of habituation typically pertains to how people feel rather than how they rate or evaluate stimuli.

Aggregation of Stimuli Expressing Emotion

Aggregating stimuli expressing emotions can help people to navigate social interactions, and there is a vast literature on how people encode and perceive such stimuli (Elfenbein & Ambady, 2003; Elfenbein, Marsh, & Ambady, 2002). However, most prior research has focused on the perception of individual pieces of emotion information presented as momentary stimuli (Goldenberg, Schöne, et al., 2022). One related concept that has been studied is ensemble coding, which explores how individuals combine multiple instances of stimuli to form summary statistics —typically an average (Whitney & Yamanashi Leib, 2018). There are a few studies in the ensemble coding literature that did focus emotion expression of multiple faces. While the earlier research found that people are able to accurately assess the mean emotion expressed by a crowd (Haberman et al., 2009; Haberman & Whitney, 2009), more recent work on multiple facial expressions presented either together as a crowd or sequentially over time found that people systematically overestimate multiple expressions. Specifically, Goldenberg, Weisz, et al. (2020) found that people tend to overestimate the average emotions expressed by crowds, especially larger ones. In a separate study, Goldenberg, LaFollette, et al. (2022) found that this effect is even stronger when crowds contain more Black individuals expressing negative emotions. A study most closely aligns with the goals of this research investigated how people estimate the average emotions expressed in sequences of faces (Goldenberg, Schöne, et al., 2022). Participants in this study were shown sequences of faces varying in length from 1 to 12, and they were asked to rate the average emotionality of the faces within each sequence on a scale of 0 (neutral) to 50

(intense emotion). The facial expressions were generated by interpolating 48 morphs between a neutral face and the corresponding angry and happy expressions, with each increment representing a 2% change on a 1-50 scale. These ratings were then used to calculate the objective mean. The findings revealed that participants consistently overestimated the average emotional expression of the faces in the sequences, especially for longer sequences. These studies suggest that people generally tend to overestimate emotionality when faced with multiple expressions presented either simultaneously or sequentially.

One major limitation of these overestimation findings is that they were based on the perception of faces, which are processed differently than written words (Hagen et al., 2021). While faces rely on automatic visual processing, affect in text requires both visual processing and semantic processing to decode the meaning of the words in a sentence (Derks, Fischer, & Bos, 2008; Harris & Paradice, 2007), resulting in a potentially deeper level of processing. Other research investigating the aggregation of purely visual features, including two studies that examined people's ability to evaluate the average size of circles appearing in a sequence, provides consistent evidence for a general overestimation of affective intensity (Albrecht & Scholl, 2010; Hubert-Wallander & Boynton, 2015). In contrast, the only study looking into the sequential perception of texts (Brady et al., 2022) found evidence of overestimation only in one of the two conditions, which raises the general question of whether overestimation in aggregation is restricted to perceptions based on purely visual features.

How do people arrive at overestimations of affective information in sequences? People have a tendency to pay relatively more attention to affective stimuli, including faces and words (Eimer & Holmes, 2007; Kissler, Herbert, Peyk, & Junghofer, 2007), especially when their affective content is negative (negativity bias, Baumeister et al., 2001). The aforementioned studies, which investigated sequences and crowds of faces, offer preliminary

evidence that supports the important role of attention. Goldenberg, Weisz, et al. (2020) conducted an eye-tracker study to examine the estimation of emotion expressions in crowds, revealing that participants tended to fixate their gaze for extended periods on emotional faces as opposed to neutral faces. Furthermore, in another study examining sequences of faces, Goldenberg, Schöne, et al. (2022) demonstrated that increasing the salience of emotional faces by adding a red square around them resulted in an amplified overestimation effect. Conversely, adding a red square around neutral faces decreased the overall average rating. This increased attention potentially brings into play two mechanisms that could lead to overestimation when calculating aggregated affect from a sequence of texts.

First, people may not be able to remember every item in the sequence by the time they have to estimate its average (Tong, Dubé, & Sekuler, 2019). Research on working memory suggests that people are able to store around four items in visual working memory (Luck & Vogel, 1997; Wilken & Ma, 2004), and intense affective texts are more likely to occupy those limited memory slots (Adelman & Estes, 2013; Kensinger & Corkin, 2003). In support of the mechanism, the study focusing on sequences of faces found that people better remembered faces that expressed stronger emotions (Goldenberg, Schöne, et al., 2022). Additionally, they also found that the order influenced the degree of overestimation, with faces presented later in the sequence—being more recent in memory—exerting a greater impact on the overall rating compared to earlier faces. Taken together, these two findings in the sequences of faces would suggest the important role of memory in overestimation.

Alternatively, even if people can remember each item, they may assign more weight to the more emotionally intense ones, as extreme information is perceived as more diagnostic or representative (Fiske, 1980). When people have to assess an event or a person, they are more influenced by extreme (e.g., uncommon or intense) information (Skowronski &

Carlston, 1989), especially when that information is negative (Skowronski & Carlston, 1987). If either of these mechanisms is the primary driving factor, another hypothesis emerges regarding the influence of sequence length. As the length of the sequence increases, there is a greater likelihood of encountering extreme stimuli that disproportionately impacts an individual's judgment of the sequence. In other words, the longer the sequence, the more pronounced the tendency to overestimate should become (if the above reasoning is correct).

Research Questions and Goals

I conducted a series of four experimental studies with the goal of evaluating the estimation of average emotional language in text responses to news articles, the relationship between sequence estimation and the rating of the individual news articles, as well as the mechanism behind the aggregation estimation. I hypothesized that attention to affective words drives the aggregation estimation (see Appendix of Chapter 5 for a version using words), similar to how attention to emotional faces drives the estimation of face sequences.

To test this hypothesis, Studies 5 - 6 were designed to detect the existence of overestimation in the evaluation of affective text sequences in response to a news article, addressing potential challenges to the main finding in each experiment. These studies had four pre-registered hypotheses (https://osf.io/y2pj5/). The first hypothesis was that participants would overestimate the intensity of the average affect expressed in a sequence of responses (amplification effect). The second hypothesis was that amplification would be stronger for longer sequences. The third hypothesis focused on the difference between sequences of positive and negative news articles and their responses. While I expected amplification to be stronger for negative affective content, I also note that previous research has found that the difference between negative and positive content is relatively small and

inconsistent (see Goldenberg et al., 2020). Finally, I hypothesized that the emotional language intensity of the news article would influence the rating of the sequence, and vice versa.

Studies 7 and 8 were designed to investigate the mechanism underlying the overestimation of emotional language in the evaluation of text sequences. My specific aim was to investigate whether increased attention to emotional words results in better memory of emotional texts, and if affective texts are perceived as more representative or diagnostic of a sequence.

Study 5: Evaluation of Emotion Expressions in Text Sequences

The goal of Study 5 is to test whether people overestimate the emotion expressed in sequences of texts produced in response to fictional news articles. In line with my pre-registration (https://osf.io/s5q3e/), I hypothesized that: (1) participants would overestimate the emotional language intensity of the sequence compared to the computed average based on the individual ratings, (2) this overestimation would increase for longer sequences, (3) negative sequences would elicit greater overestimation compared to positive ones, and (4) that overestimation should be more pronounced in emotionally intense contexts.

Method

Participants. Based on my power analysis and goal to evaluate our stimuli (all details in the Appendix of Chapter 5), I recruited 300 native English speakers from the United Kingdom and the United States via Prolific. Following my pre-registered exclusion criteria, I excluded 11 participants who failed more than three attention checks or missed survey

responses. In addition, I excluded 7 participants whose mean estimation exceeded 2.5 standard deviations from the total sample. This resulted in a final sample size of 282 participants (145 females, 135 males, 1 other, 1 who prefers not to say, age: M = 25.35, SD = 7.37, range = 18-70). Participants were compensated $4.60 for their participation in the 40-minute task.

Stimulus set. This experiment utilized two types of stimuli. The first type of stimuli consisted of 50 fictional news stories, which I developed, comprising both positive (25) and negative stories (25) that resembled local newspaper articles. My objective in creating the fictional news stories was to achieve a range of emotional language intensities while ensuring that the negative stories were not significantly different from the positive stories in terms of their intensity. The second type of stimuli was 25 responses to these news articles, varying in affective intensity.

To evaluate the emotional language intensity of the news stories and generate responses, I conducted a pilot experiment where participants first rated each news story on a 0 to 9 scale, with 0 being neutral and 9 being highly emotional (see Appendix of Chapter 5 for more details). For each of the news stories, participants were asked to write a response as if they were posting online comments about the news articles, using a randomly assigned pre-defined intensity rating ranging from 0 (neutral) to 9 (very emotional) to generate responses with varying levels of emotional language intensity.

These responses were manually edited to ensure they were no longer than 120 characters. In total, there were 20 responses per news story, with two responses for each intensity on the 0-9 scale. It's important to note that the responses were not evaluated separately for intensity, unlike the news articles, so they may not align precisely with the 0-9

scale[14]. No separate evaluation pilot was conducted for the responses since participants in this experiment rated each response in the sequence themselves to calculate the objective mean. The news articles and responses had the same valence, which means that the responses covered a range of emotional language intensities from neutral to negative or from neutral to positive, depending on the valence of the news article. The emotional language intensity of the positive news articles did not differ significantly from that of the negative ones, as detailed in the Appendix of Chapter 5.

Procedure. Participants completed 50 trials, following two practice trials, which were excluded from the analysis. At the beginning of each trial, participants were randomly presented with one of the 50 fictional news articles. After 3 seconds, a rating scale appeared and participants were asked to rate the emotional language intensity of the news article on a scale from 0 (neutral) to 100 (very emotional). I switched from the 0 – 9 Likert scale to a 0 – 100 slider to allow for more fine-grained variation and to make it more difficult for a participant to remember the exact rating that they have given to each stimulus. Next, a fixation cross was shown in the centre of the screen for a randomly determined duration of 400-600ms, after which one of the 20 possible responses was presented to the participant. Participants were asked to rate the emotional language intensity of the response on the same 1 to 100 scale, after 1 second when the scale appeared. Each trial consisted of 4-10 sequences

[14] Utilizing the average intensity ratings obtained from Study 5, I conducted a retrospective analysis to examine the correlation between the intended intensity levels and the actual ratings provided during the experiment. The analysis unveiled a significant correlation between the intended intensity levels, determined by selecting the stimuli based on the pre-assigned intensity of the pilot, and the average intensity ratings from Study 5 ($r = .79$, $t(998) = 40.94$, $p < .001$). These results suggest a close alignment between the intended intensities and the rated intensities for the selected stimuli.

of a fixation cross, response, and individual response rating. After watching and rating all of the responses, participants were asked to estimate the average emotional language of the sequence on the same scale. Each news article and response were presented only once to each participant. After completing the main task, participants again filled out a short survey (see Appendix of Chapter 5).

Measures. My main measurement variables included three types of ratings: the rating of the news article, the individual ratings of the responses, and the total average estimation of the sequence. To calculate overestimation, I compared the total average estimation to the average of the individual ratings provided by each participant. By using individual ratings for this calculation, I was able to obtain a more precise estimation of the objective mean without relying on scores from previous raters as in other previous studies.

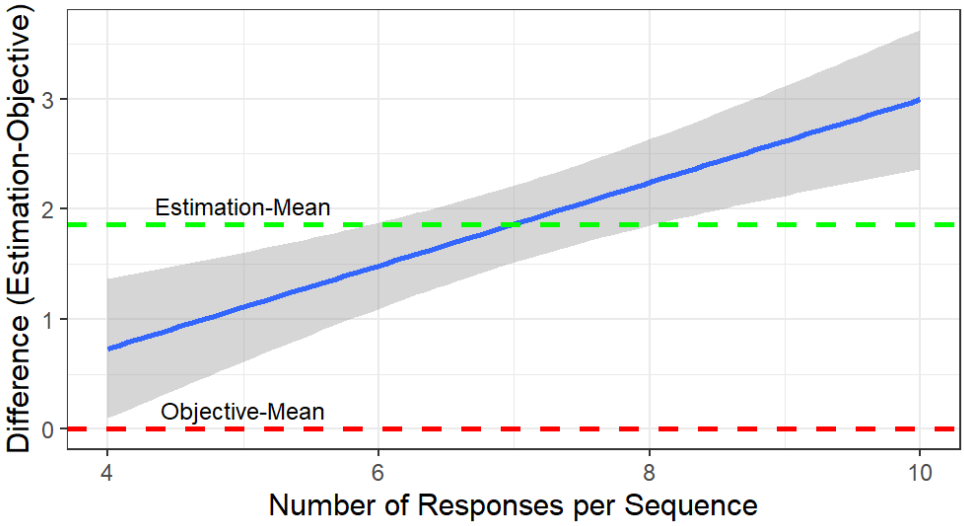
Results

As participants estimated multiple sequences, I used linear mixed models analysis for repeated measures for the following models. I utilized the packages lme4 (Bates, Mächler, Bolker, & Walker, 2014), and the lmerTest package (version 3.1.2 – Kuznetsova et al., 2017), implemented in the R programming language (version 4.1.2). To test hypotheses 1-3, I included users' news article intensity ratings as a covariate, as I suspected (hypothesis 4) that this rating might influence the sequence average rating, and I wanted to determine if there was an overestimation effect above and beyond the effect of the news article. Additionally, I included random intercepts for participants and news articles to account for individual differences in emotional language perception and news article intensities, respectively.

Overestimation of the average emotional language expressed. To evaluate whether participants overestimate the average affect expressed in a sequence of texts, I compared the difference in emotional expression values between two types of means: The first was the actual sequence mean, calculated from individual response ratings provided by each participant throughout the sequence. The second was the estimated mean of the sequence, reported by the participant. In line with my first hypothesis, I found that participants significantly overestimated the objective mean of the sequences compared to the aggregated mean of their individual ratings even after controlling for the news article's intensity (b = 1.86 [1.27, 2.45], SE = 0.29, t (8121.17) = 6.21, $p < .001$, $R^2 = .37$).

Overestimation of the average emotional language expressed by sequence length. Next, I wanted to find out if the sequence length influenced the estimation of the mean. For this analysis, my dependent variable was calculated as the difference between the participant's estimation and the mean emotional language of the sequence. My independent variable was the number of texts (ranging from 4 to 10) in a given sequence, after controlling for the intensity of the news article. I found that increasing the sequence length was associated with a greater degree of overestimation (b = 0.38 [0.21, 0.55], SE = 0.085, t (4081.98) = 4.43, $p < .001$, $R^2 = .14$, see Figure 5.1).

Figure 5.1 *The comparison between estimated and actual average ratings of the sequence.*



Notes. *Analysis was done using a linear mixed model. The x-axis represents the number of responses in the sequence. The y-axis represents the difference between participants' estimation of the average sequence and the actual average. Data are presented as mean value +/- confidence intervals. The green dotted line represents the average amplification. The red dotted line represents the actual emotion expression of the response's mean.*

Overestimation of the average emotional language expressed by valence. To assess whether participants' overestimations differed between positive and negative sequences, I again used the difference between participants' estimations and the mean word affect as the dependent variable. The independent variable was valence, specifically whether the sequence has a positive valence or negative valence. I found that participants' overestimation of the mean more did not differ significantly between negative and positive sequences (b = -0.81 [-1.92, 0.30], SE = 0.56, t (43.02) = -1.43, p = .16, R^2 = .14).

Overestimation of the average expressed emotional language by news article intensity. Finally, to test my fourth hypothesis, I wanted to test if the emotional language intensity of the context (news article) influenced the overestimation of the sequence. I again used the estimation difference as described in the previous analysis as my dependent variable. However, instead of using the intensity of the news article rating as a covariate, I treated it as the main independent variable. Results indicated that the intensity rating of the news article was positively associated with overestimation (b = 0.065 [0.05, 0.08], SE = 0.0073, t (1836.57) = 8.86, $p < .001$, $R^2 = .14$).

To confirm that participants based their estimation of the mean on the responses rather than the news article, I conducted additional analysis predicting participants' estimation using both the news article rating and the objective mean. The standardized coefficient for the objective mean (b = 0.98 [0.95, 1.00], SE = 0.013, t (3444.46) = 73.52, $p <$ .001, $\eta^2 = .62$, $R^2 = .71$) was significantly larger than that for the news article rating (b = 0.069 [0.95, 1.00], SE = 0.0079, t (2139.70) = 8.77, $p < .001$, $\eta^2 = .03$, $R^2 = .71$), indicating that participants' estimation was more influenced by the responses than the context. The chi-square test confirmed this result ($\chi 2$ (1) = 2623.2, p < .001).

Discussion

Findings from Study 5 indicated that people tended to overestimate the average emotional language in texts. In line with the second hypothesis, overestimation increased with longer sequences. However, I did not find significant differences between positive and negative sequences. I also found that the context of the texts played a role in shaping people's overestimation tendencies. Specifically, more intense contexts, in this case news articles led to a greater degree of overestimation. This raises the question of whether overestimation is

caused solely by the intensity of the news article. My initial analysis suggested that while the sequence rating had a relatively stronger influence on sequence estimation, the effect of news article intensity appeared to be significant, which could potentially be enough to lead to overestimation. Second, from a practical perspective, users of social media platforms may encounter other people's responses to an article before reading the article themselves, which raises the question of whether they may overestimate the sequence even before encountering the news article. Finally, in a similar vein, it is unclear if the estimation of the sequence would then affect how intensely people perceive the context of the texts. In other words, it remains unclear whether this relationship between context and overestimation is bi-directional. I explore these possibilities in the next experiment.

Study 6: Analysing the Effect of the Average Emotion Expression Estimation of the Responses on the Perception of the Corresponding News Article

Having established that people tend to overestimate the emotional intensity of sequences of textual responses to fictional news articles, my aim in Study 6 was to investigate whether this overestimation effect persists when the news article (i.e., context) is presented after the sequence has been rated. I also aimed to investigate whether people's judgments of the overall sequence influence their subsequent rating of the news article. Additionally, to explore how the perceptions of affective content shape our online behaviour, I examined how participants' evaluations of the news article and responses affect their hypothetical willingness to share the news article on social media. Specifically, I hypothesized that (1) participants would overestimate the emotional language intensity of the sequence compared to the computed average based on individual ratings, (2) this overestimation effect would increase for longer sequences, and (3) negative sequences would elicit greater overestimation

than positive ones. Furthermore, I hypothesized that (4) overestimating the sequence would increase the emotional language intensity rating of the news article. In an exploratory analysis, I also wanted to examine the extent to which participants' evaluation of the news article and the sequence of responses influenced their hypothetical willingness to share the news article on social media.

Method

Participants. I recruited 150 native English speakers from the United Kingdom and the United States using Prolific, which exceeds the number suggested by my power analysis based on Study 5 (see Appendix of Chapter 5). However, given the uncertainty about whether the amplification effect would be reduced when the news article was presented at the end of the trial instead of the beginning, I opted to increase the number of participants recruited. I excluded 8 participants who failed more than 3 attention checks or had missing data. Additionally, following my pre-registered exclusion criterion that the mean estimation of any participant should not be more than 2.5 standard deviations away from the sample mean, I excluded 2 more participants. My final sample consisted of 139 participants (70 females, 67 males, 1 other, 1 who preferred not to say, age: M = 24.68, SD = 6.79, range = 18-57). Participants were compensated $4.60 for their participation in the 40-minute task.

Stimulus set. In this experiment, I again used the two types of stimuli created for Study 5, namely, the fictional news articles that were manually developed and their corresponding responses. I leveraged the ratings made by participants in Study 5, to further evaluate and select both types of stimuli for use in Study 6. This selection process aimed to make the positive and negative stimuli as similar as possible in terms of their emotional language intensity. This was done to improve our ability to detect differences between the

stimuli, which might have been obscured by differences in the distribution of emotional language intensities. In Study 5, 50 news stories were rated on a scale from 0 (neutral) to 100 (very emotional). I selected 40 news stories (20 positive and 20 negative stories) for Study 6 based on two criteria: (1) that the mean emotional language ratings of the positive and negative stories were not significantly different (positive: M = 63.24, SD = 10.7; negative: M = 64.12, SD = 13.5, t (36.15) = 0.22, p = .79); and (2) that the emotional language intensity ratings for both positive (W = 0.90, p = .053) and negative (W = 0.97, p = .85) stories were normally distributed according to Shapiro-Wilk tests. Of the 20 responses, I selected 15 responses for each of the 40 remaining news stories such that the mean of all positive responses (M = 53.09, SD = 17.01) was not significantly different from the mean of the negative responses (M = 54.79, SD = 15.20, t (590.15) = 1.29, p = .19).

Procedure. For Study 6, I utilized the same three-part task structure as in the previous experiment, with a few modifications. Participants completed 15 trials, following two practice trials that were excluded from the analysis. In contrast to the previous experiment, participants were presented with a sequence of 4-10 responses to one of the 40 news articles first, instead of seeing the news article at the beginning of the sequence. Again, participants rated each of these individual responses on a scale from 0 (neutral) to 100 (very emotional). The responses were randomly selected from the 15 responses corresponding to one of the 40 fictional news stories. A fixation cross appeared at the centre of the screen for 400-600ms between each response. After viewing the sequence of responses, participants were asked to rate the average emotional language intensity of the responses (sequence average estimation). Next, they were shown the news article that corresponded to the sequence of responses they had just seen and were asked to rate its emotionality using the same scale as described above. As a final modification, participants were asked to indicate the likelihood of sharing the news

story on social media using a scale ranging from 0 (very unlikely) to 100 (very likely) at the end of each trial. After completing the primary task, participants were directed to a brief survey (see Appendix of Chapter 5).

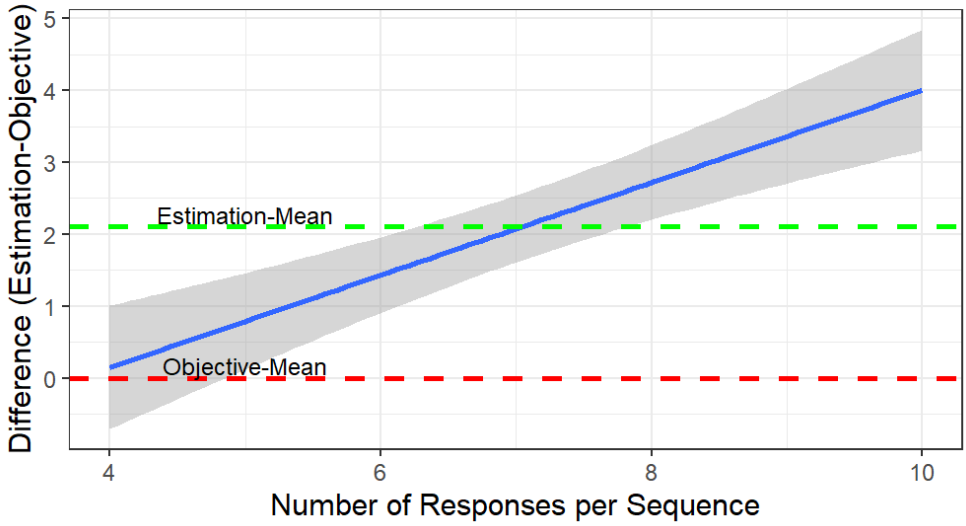
Results

The analysis plan, including the models and statistical packages used, was the same as in Study 5 with one exception. Since the context of the news article was only introduced at the end of the trial, the emotional language intensity rating of the news article was not included as a covariate in the analysis. However, as in Study 5, I used individual participants and news stories as random intercepts in my models to adjust for individual rating differences and stimulus effects, respectively.

Overestimation of the average emotion expressed. Starting with my first hypothesis, I replicated the finding that participants tended to overestimate the average emotional language expressed by the sequence compared to the objective mean (b = 2.10 [1.35, 2.86], SE = 0.38, t (3968.97) = 5.46, $p < .001$, R^2 = .36), despite the presentation of the news story context after the sequence rating.

Overestimation of the average emotion expressed by sequence length. Supporting my second hypothesis, I found that longer sequence lengths were associated with greater overestimation (b = 0.63 [0.40, 0.86], SE = 0.11, t (2049.88) = 5.32, $p < .001$, R^2 = .10, see Figure 5.2), further replicating the results of Study 5.

Figure 5.2 *The comparison between estimated and actual average ratings of the sequence.*



Notes. *Analysis was done using a linear mixed model. The x-axis represents the number of responses in the sequence. The y-axis represents the difference between participants' estimation of the average sequence and the actual average. Data are presented as mean value +/- confidence intervals. The green dotted line represents the average amplification. The red dotted line represents the actual emotion expression of the response's mean.*

Overestimation of the average emotion expressed by valence. After making the positive and negative stimuli more similar in this experiment, I repeated the analysis testing whether participants overestimated to a different extent for positive and negative sequences. However, I again found no difference in overestimation for negative compared to positive sequences (b = 0.022 [-1.18, 1.23], SE = 0.61, t (37.62) = -0.036, p = .97, R^2 = .09).

Influence of overestimation of emotional language in sequence on news article intensity. Next, I examined whether the overestimation of the mean emotional language in the sequence affected the intensity rating of the news article. The emotional language

intensity rating was the dependent variable, while the difference between estimations and mean response emotional language was the independent variable. Our findings showed a positive association between the extent of participants' overestimation of the average emotional language expressed in the response sequences and the perceived intensity of the news article (b = 0.19 [0.11 – 0.27], SE = 0.040, t (1964.00) = 4.67, $p < .001$, R^2 = .45).

Influence of emotional language estimation of response sequence and news article on sharing tendency. Finally, I wanted to test how both emotional language evaluations contribute to a participant's hypothetical sharing willingness. I used two independent variables: the participant's estimation of the average emotional language in the responses, and the emotional language intensity rating of the news article. My results revealed that both intensities of the sequence estimation (b = 0.073 [0.011, 0.13], SE = 0.031, t (2018.57) = 2.34, p = .02, η^2 = .0027, R^2 = .59), and the intensity rating of the news article made unique contributions to the hypothetical sharing tendency (b = 0.56 [0.51, 0.61], SE = 0.023, t (2023.10) = 23.38, $p < .001$, η^2 = .21, R^2 = .59). The emotional language intensity of the news article had a larger impact on the sharing behaviour as indicated by the standardized coefficients as well as confirmatory chi-square test ($\chi 2$ (1) = 118.4, $p < .001$).

Discussion

To summarize the findings of Study 6, first, I showed that people tend to overestimate the average emotion expressed in responses to news articles, even when the news article is presented after the rating. Second, I replicated the finding that an increase in sequence length is associated with an increase in overestimation. As in Study 5, I could find no evidence that this overestimation differed between positive and negative trials, despite making both types of stimuli more similar in their affective intensity. Additionally, I demonstrated that

overestimation influences the way people judge the emotional language intensity of news articles (i.e., context). Furthermore, I provided initial evidence that the reported intention to share a news article depends not only on the emotional language intensity of the news article but also on the estimation of the average emotional language of the response sequence.

After demonstrating bi-directional influences between sequence and context estimations and their contribution to online sharing behaviour, the question arises of what mechanisms underlie these estimations. Specifically, why do people tend to overestimate the emotional language intensity of response sequences to news articles? In the next studies, I explore potential explanations for this amplification.

Studies 7 - 8: Evaluating Possible Explanations for the Overestimation of Emotion Expressions in Text Sequences

Having established that individuals tend to overestimate the emotional language intensity of response text sequences to fictional news articles, the question arises as to why this phenomenon occurs. There are multiple potential mechanisms that could explain how people arrive at an exaggerated average. One possibility is that individuals may not remember all items in a sequence with an equal likelihood when making their assessment. As previously shown in studies of emotional faces (Goldenberg, Schöne, et al., 2022), if individuals have stronger memory for more emotionally intense items, this could lead to an overestimation of the emotional average. Alternatively, individuals may assign greater weight to more intense perceptions when making their evaluation, even if they remember all items equally well. The perception that more intense expressions are more normative could be one reason why individuals assign increased weight to the estimation of the average expression, thereby contributing to the overestimation of the sequence's average expression. In a study that did

find that people have a tendency to perceive more intense opinions as more typical or representative, Goldenberg, Abruzzo, et al. (2022) speculated that this tendency might be due to peoples' preference for coherent opinions, with the presence of strong opinions serving as an indication of such coherence.

In these studies, I aimed to investigate two potential mechanisms that potentially contribute to the overestimation of emotional language averages, in addition to the previously tested hypotheses (1-4). Specifically, I examined whether individuals (6) remember more intense texts better and whether they (7) perceive more affective texts as more representative of others' feelings about the news story. To test both mechanisms, I developed two separate experimental studies. Study 7 focused on Hypothesis 6, examining whether people tend to remember more intense emotional expressions better. In contrast, Study 8 tested Hypothesis 7, exploring whether more intense responses are perceived as more representative.

Method

Participants. I recruited 150 native English speakers from the United Kingdom and the United States through Prolific for both studies. I recruited more participants than suggested by the power analysis based on Study 6, as I did not vary the sequence length in the design of these studies (see Appendix of Chapter 5 for more details). Previous studies that have used sequences of faces have shown that the amplification effect is smaller when the number of stimuli does not vary between sequences (Goldenberg, Schöne, et al., 2022).

In Study 7, I excluded 10 participants who failed more than 3 attention checks or had missing data. Additionally, I excluded 2 participants in line with my pre-registered exclusion criteria that the mean estimation should not be more than 2.5 standard deviations away from the sample mean. My final sample consisted of 138 participants (68 females, 70 males, age:

M = 39.81, SD = 13.73, range = 18-74), who received a compensation of $5.20 for their participation in the 45-minute task.

In Study 8, I removed 4 participants whose mean estimation were more than 2.5 standard deviations away from the sample mean in line with my pre-registered exclusion criteria. The final sample comprised 146 participants (75 females, 71 males, age: *M* = 41.95, *SD* = 14.90, range = 20 -77). All participants completed a 50-minute task and received compensation of $5.75.

Stimulus set. For this experiment, I used the same stimuli as in Study 6, in which I ensured that the emotional intensity ratings were statistically indistinguishable. The stimuli consisted of 20 positive and 20 negative news articles, each with 15 corresponding responses per article.

Procedure. Both studies were divided into two blocks, with the first block designed to replicate previous findings, and the second block intended to test one of the two potential mechanisms. A short survey was administered to participants after the completion of the task. To ensure that participants understood the instructions, they began each task with two practice trials per block. Once they had completed these practice trials, they moved on to the actual experiment, which consisted of 30 trials in total (15 for each block). During the first block, participants completed 15 trials that closely resembled those in Study 5, with the exception that they were presented with 8 responses per sequence, rather than the 4-10 responses used in earlier studies. To avoid introducing the effects of sequence length on the likelihood of remembering items, I decided to fix the number of responses per sequence to 8. This decision was also informed by the fact that participants in a previous experiment (Goldenberg, Schöne, et al., 2022) saw the same sequences in the memory task block. As in both previous experiments, participants provide four different types of ratings in the

following order: (1) news article emotion intensity rating, (2) 8 individual response emotion intensity ratings (the basis for computing objective mean), (3) sequence average emotion response rating, and (4) hypothetical willingness to share the news article on social media.

In Study 7, during each trial of the memory task block, participants viewed three text responses to a news article. One of these responses was identical to a response presented in the previous block, while the other two responses had not been shown to participants previously. The order in which the correct and the false memory items were presented was randomized. Participants were asked to select the response which had been previously presented in the sequence that they had seen.

In Study 8, during the representativeness judgment task block, unlike in the memory block of Study 7, participants first saw the fictional news article. After that, three text responses to the news article were shown above each other on the screen. All of these responses were randomly selected from the sequence that participants saw during the previous block. The order in which the items were presented was randomized. Participants were then asked to select which of the three responses was most representative of the emotion that others would express in response to this article by clicking on it. I decided to phrase this question in terms of “representativeness” instead of “being normative”, as I suspected that participants would understand this more intuitively.

Results

The statistical analysis for testing the hypothesis related to overestimation, including the models, covariates, and random effects, was identical to that used in Study 5. However, due to the fixed number of responses (8), I was unable to test hypothesis 2, which explored the effect of sequence length on overestimation. To examine the association between

emotional language intensity and memory as well as representativeness (hypotheses 6-7), I used a mixed-effects logistic regression model, since selecting a response or not selecting it is a dichotomous variable. In Study 7, I also included the length of the response in characters (M = 61.02, SD = 34.54) as a covariate, as it could potentially influence the likelihood of remembering the item.

Overestimation of the average emotional language expressed. In both Studies 7 - 8, I found that participants significantly overestimated the objective mean of the sequences compared to the aggregated mean of their individual ratings (Study 7, b = 1.15 [0.29, 2.01], SE = 0.43, t (8121.17) = 6.21, p = .0087, R^2 = .54; Study 8, b = 2.12 [1.36, 2.88], SE = 0.38, t (4179.49) = 5.48, p < .001, R^2 = .53).

Overestimation of the average emotional language expressed by valence. My results showed no significant difference in overestimation of the mean between negative and positive sequences in either of the studies (Study 7, b = -0.59 [-2.41, 1.24], SE = 0.56, t (37.86) = -0.63, p = .52, R^2 = .14; Study 8, b = 0.18 [-1.07, 1.43], SE = 0.63, t (35.99) = 0.28, p = .78, R^2 = .24).

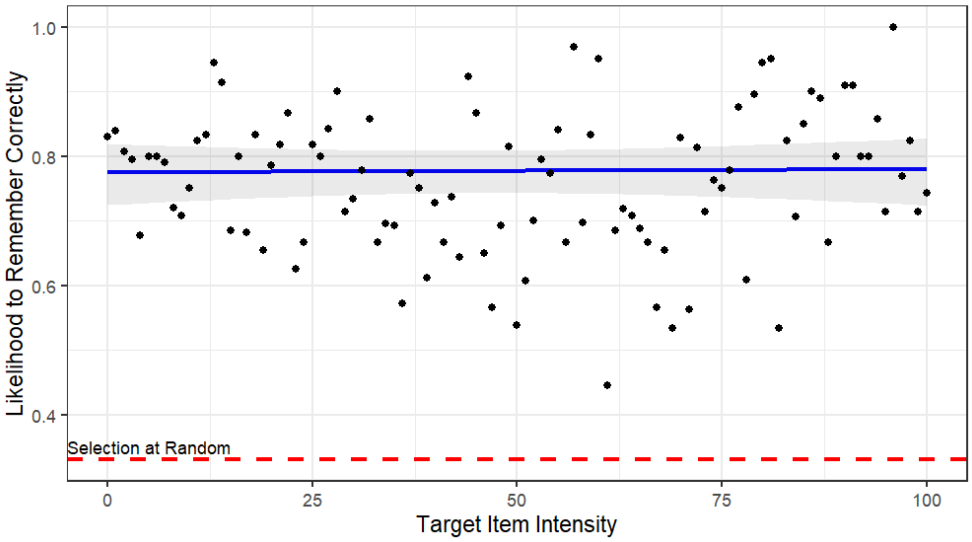
Overestimation of the average expressed emotion by news article intensity. My results revealed a positive association between the intensity rating of the news article and the level of overestimation in both studies (Study 7, b = 0.19 [0.11, 0.27], SE = 0.0073, t (1979.28) = 4.52, p < .001, R^2 = .56; Study 8, b = 0.20 [0.13, 0.28], SE = 0.039, t (2083.51) = 5.10, p < .001, R^2 = .59).

Influence of emotional language estimation of response sequence and news article on sharing tendency. In both studies, my findings revealed that both variables contributed to the tendency to share (Study 7, estimation: b = 0.15 [0.10 – 0.20], SE = 0.024, t (2032.71) = 6.03, p < .001, η2 = .02, R^2 = .61; news article intensity: b = 0.37 [0.51, 0.61], SE

= 0.019, *t* (1617.52) = 19.01, $p < .001$, $\eta 2 = .18$, $R^2 = .61$; Study 8, estimation: *b* = 0.17 [0.13, 0.22], *SE* = 0.024, *t* (2161.21) = 7.20, $p < .001$, $\eta 2 = .02$, $R^2 = .61$; news article intensity: *b* = 0.37 [0.33, 0.40], *SE* = 0.019, *t* (1761.73) = 19.96, $p < .001$, $\eta 2 = .18$, $R^2 = .61$). However, the emotional language intensity of the news article had a greater impact on sharing behaviour in both studies, as confirmed by standardized coefficients and a chi-square test (Study 7, $\chi 2$ (1) = 37.12, $p < .001$; Study 8, $\chi 2$ (1) = 29.42, $p < .001$).

Association between emotional language intensity of response and memory. In Study 7, I aimed to examine if people have a higher likelihood of recalling emotionally intense texts, meaning, that if a participant tends to remember more intense texts, they may also overestimate the emotional language of a sequence more. I used the correctness of the judgment of whether a text was in the sequence as my dependent variable and the participant's emotional language intensity rating of the target stimuli (correct text response) from the sequential text amplification task trials as an independent variable. To accomplish this, I created a long-format dataset, where each response in a trial was an individual row. Study 7 did not find any evidence that participants were more likely to remember responses with higher emotional language intensity (*b* = 0.00028 [-0.02, 0.13], *SE* = 0.0021, *z* = 0.13, *p* = .89, $R^2 = .59$, see Figure 5.3).

Figure 5.3 *The likelihood to remember a response based on its emotional language intensity.*

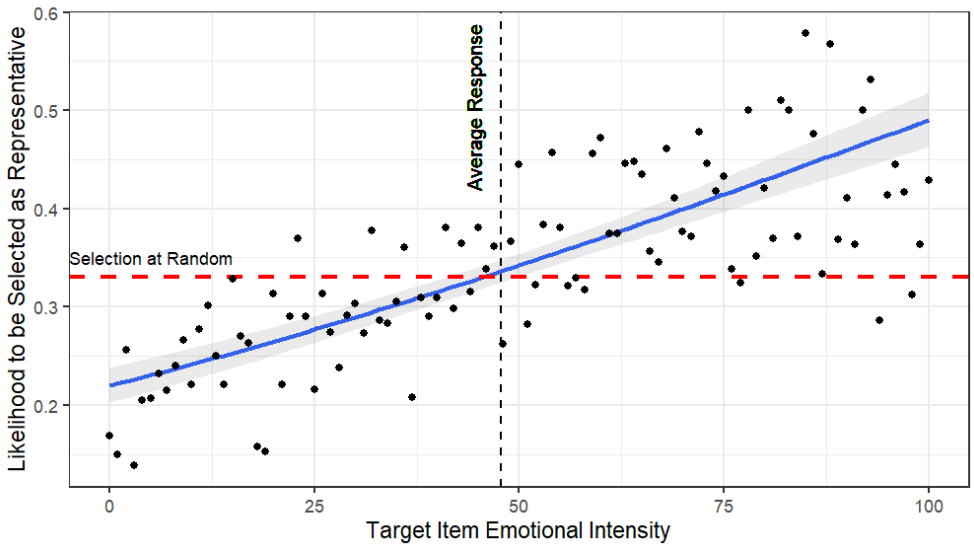


Notes. *Analysis was done using a logistic mixed model. The x-axis represents the emotional language intensity of the target stimuli as rated by the participant during the sequential text emotion estimation task. The y-axis represents the likelihood of the target response being correctly remembered. Data are presented as mean value +/- confidence intervals. The red dotted line represents selection likelihood by random chance (1/3).*

Association between emotional language intensity of a response and norm perception. Finally, in Study 8, I aimed to examine whether participants perceive emotionally intense texts as more representative, and if so, whether this perception is associated with overestimating the average emotional language expressed in response sequences. In my analysis, I calculated the probability of a response being selected based on its emotional language intensity rating. For this analysis, we again transformed the dataset into long-format. The dependent variable was whether the response was selected or not, while the independent variable was the emotional language intensity rating of the response as rated by the participant

during the sequential text amplification task. I found that participants perceived emotionally intense responses as more normative or representative of how other people would respond to the news article. (b = 0.012 [0.010, 0.014], SE = 0.017, z = 12.79, p < .001, R^2 = .035, see Figure 5.4).

Figure 5.4 *The likelihood to select a response as most representative based on its emotional language intensity.*



Notes. *Analysis was done using a logistic mixed model. The x-axis represents the emotional language intensity of the target stimuli as rated by the participant during the sequential text emotion estimation task. The y-axis represents the likelihood of the target being selected as the most representative of the sequence. Data are presented as mean value +/- confidence intervals. The red dotted line represents selection likelihood by random chance (1/3). The vertical black dotted line indicates the average of all response ratings, which could be seen as the objective most normative selection.*

Discussion

In Studies 7 and 8, the focus was on replicating previous findings and investigating the mechanism behind why people overestimate the emotional language expressed in sequences of responses. Despite successfully replicating the overestimation effect in both studies, Study 7 did not find any evidence that an increased likelihood of remembering emotionally intense responses was associated with overestimation. It seems that participants were relatively accurate in remembering the responses shown to them, but still overestimated the average affective intensity of the response sequence.

In Study 8, I investigated why people might overestimate despite not having a systematic tendency to primarily remember more intense responses. I found evidence that individuals perceive emotionally intense texts as more normative or representative of how other people would have responded to the news article. As a result, they may assign greater weight to these emotionally intense responses in their aggregation, leading to an overestimation of the average compared to their objective mean. However, so far there is no direct evidence explaining why people perceive more intense opinions or expressions as more representative. Aside from the speculation raised by Goldenberg, Abruzzo, et al. (2022) that people perceive stronger opinions as more representative because of coherence, there are other potential explanations. For example, people might work from an assumption that people tend to post unduly extreme emotional responses on social media, leading them to perceive strong opinions as more indicative of typical social media behaviour.

Overall, these studies provide evidence that people tend to overestimate the emotional language of sequences of text responses to fictional news articles. This overestimation does not seem to be due to them remembering more intense responses but may rather be due to

them overweighting emotionally intense texts as they are perceived as more normative or informative.

General Discussion

In the experiments reported in this chapter, I found that people seem to overestimate the emotional language intensity of text responses to fictional news articles. This overestimation was stronger for longer sequences. When analysing how the context of text responses influences the rating of both, I found that the rating of the context influences the rating of the sequence and vice versa, which can increase the perceived emotional language intensity of these estimations. People have a greater tendency to share affective content that they perceive to express stronger emotions. When analysing the potential explanations for this effect, I found evidence that participants had no systematic bias towards better-remembering responses that use more emotional intense language, however, they may have weighted the more intense responses into their ratings because they were perceived as more representative or diagnostic of the overall sequence.

These findings contribute to the existing literature in two significant ways. First, this set of four experimental studies expands our understanding of how social media usage, and specifically viewing affective content sequentially, can amplify our perception of events, opinions, and emotions (Tucker et al., 2018). Second, it adds to the existing body of research on ensemble coding, specifically the aggregation of emotional stimuli presented in sequences (Goldenberg, Sweeny, Shpigel, & J. J. Gross, 2020; Haberman et al., 2009; Haberman & Whitney, 2009). In this case, these experiments offer potential explanations of how the aggregation of average emotional language is achieved.

Implications of the Findings.

These findings contribute to the wide array of studies suggesting that the design of social media can act as a magnifying glass for emotions and opinions (Brady et al., 2019), resulting in potential harm for its users. This amplification might be the result of the interplay of human tendencies exhibited by perceiver and their networks, which can be further perpetuated by a platform's algorithm. One of the reasons for the overestimation of affective intensity lies within the perceiver themselves, as highlighted in Chapters 2 and 3. These chapters presented evidence of users' inclinations to engage with negative affective content and content involving intergroup conflict, both of which contribute to a decline in emotional well-being and an increase in political polarization (de Mello et al., 2022). Moreover, as suggested by Brady et al. (2022), viewers perceive emotional language intensity as stronger than intended by the content creator. In addition to these two established perceiver-based processes that are amplified by social media, the current set of studies suggests that sequential presentation can be another factor amplifying the estimation of affective content online, which may further contribute to the harmful effects of social media.

Another factor contributing to the magnifying effect of social media is the network of users that the perceiver follows. Previous research has suggested that the network is more likely reinforces the production of intense affective content on social media by both liking and sharing others' affective content (Brady et al., 2021; Frimer et al., 2023). Additionally, the findings presented in this chapter shed light on how the content generated by user networks directly impacts how individuals evaluate news articles. Specifically, if users encounter multiple instances of affective content before viewing a news article, they tend to perceive the intensity of the news article as higher than if they had not been exposed to the sequence. Furthermore, if users come across a news article that employs intense emotional

language, they also overestimate the average sentiment expressed subsequently by other users. This feed-like design inherent in social media platforms fosters a misperception of both the network itself and the content being discussed, adding another layer to our understanding of the potential harm that social media can inflict on its users.

Finally, in this chapter, I presented direct evidence that an increase in the perceived intensity of affective content leads to a higher intention to share it. When combined with the reinforcing nature of user networks mentioned earlier, it becomes evident that users are more inclined to engage with more intense affective content. In light of social media platforms' vested interest in sustaining user engagement, it is plausible that their algorithms are intentionally designed to prioritize the promotion of content that aligns with users' past interactions (Brady et al., 2023). As a result, due to perceivers' inclination to engage with emotionally charged content, especially negative affective content (Rathje et al., 2022), and the perpetuating dynamics of the social networks and the platform's algorithm, social media platforms can inadvertently contribute to a distorted and potentially detrimental appraisal of emotions, events, and opinions. Within the scope of these particular experiments, I found no evidence to support the notion that people, due to the negativity bias, pay more attention to and consequently overestimate negative sequences more than positive sequences. In contrast, my findings indicate that people overestimate positive and negative sequences equally. Nevertheless, I argue that the misperception of both positive and negative emotional language intensity can be maladaptive, as individuals may still misconstrue affective cues, thereby impacting their overall emotional judgment.

The research presented in this chapter also contributes to the ensemble coding literature concerning the process of aggregation. It aligns with previous research on the overestimation of emotional facial expressions in sequences and crowds (Goldenberg,

LaFollette, et al., 2022; Goldenberg, Schöne, et al., 2022; Goldenberg, Weisz, et al., 2020), indicating that the tendency to overestimate emotional expressions is not limited to facial cues alone but can extend to the perception of affect conveyed through texts and words (see Appendix of Chapter 5). It is worth noting that this phenomenon has thus far only been observed in a single study that specifically focused on expressions of high-moral outrage. This demonstration of a general and consistent overestimation in the text is particularly intriguing, as it suggests that amplification can occur in deeper processing involving both visual and semantic aspects, going beyond mere visual characterizations of faces or objects.

The findings presented in this chapter indicate a novel mechanism to elucidate the effects of sequential amplification. They suggest that the overestimation of affective content in texts is not solely a result of biased memory, as observed in facial expression sequences (Goldenberg, Schöne, et al., 2022). Instead, the overestimation may also originate from individuals attributing greater significance to texts that convey stronger emotions. Such emotionally charged expressions are perceived as more normative or diagnostic by individuals, prompting them to assign greater weight to such content. This mechanism provides a new perspective on how the sequential presentation of affective content can lead to amplification effects. Further research is needed to clarify why and in what contexts people see more intense emotional expressions as more normative.

Limitations and Future Directions.

The experiments presented in this chapter have shed light on a remarkable proclivity for amplification—a phenomenon of utmost significance concerning social media usage and the consolidation of affective content in written expression. Although the findings are intriguing, and considerable efforts were made to address alternative explanations, it is

important to recognize and address the limitations of these studies in future research endeavours. Furthermore, these experiments have engendered a host of unresolved inquiries pertaining to the tangible manifestation of amplification within authentic user interactions on real-world social media platforms, transcending the restricted boundaries of controlled laboratory environments.

One notable limitation of these experiments is their primary focus on only one of the three key actors involved: the perceiver. The role of both the network and the platform's algorithm was not explored. For example, the experiments included no information about the content's producer and how their identity might influence how users estimate the affective intensity conveyed by news articles and responses to them. In Chapter 2, I highlighted the significant influence that content producers exert in determining the content that is shared, which suggests their potential influence on our perception of affective content as well. To address this limitation in future studies, researchers could incorporate additional information about the content producer and the news article's source to gain a comprehensive understanding of how these factors may have shaped the findings and their implications.

It is crucial to acknowledge that, despite our efforts to simulate social media posts, the experimental design falls short of fully replicating a social media feed. Notably, the news articles and the corresponding responses do not appear on the same page due to the study's paradigm, which required individual ratings after participants were exposed to a stimulus. This raises the question of whether this overestimation effect can be observed in a more realistic mock social media environment and whether this exposure has any impact on the well-being of observers. These questions should be addressed in future research.

It is important to emphasise that the content presented to participants in this study was entirely randomized, devoid of any influence from users' engagement or ratings. This meant

that people could not select certain producers whose content they would like to see again. Prior research has indicated that individuals tend to gravitate towards networks that exhibit more extreme opinions, a phenomenon known as acrophily (Goldenberg, Abruzzo, et al., 2022; Goldenberg et al., 2023). Considering the presence of the amplification effect, it would be intriguing to explore how acrophily and amplification interact to contribute to the overestimation of affective content.

In a similar vein, to better emulate the role of algorithms employed by social media platforms, it would be valuable for future research to explore the impact of the presented content and its sequencing on users' judgment dynamics. By manipulating the sequence in which content is presented to users, researchers can assess how this temporal variation influences changes in their judgments over time. Drawing on the findings of Goldenberg et al. (2022), which studied order variations in sequences of faces, it can be speculated that if more intense responses are placed later in a sequence, individuals might overestimate the intensity of the news article, as well as the overall sequence, due to the influence of the recency effect. Such an investigation would provide deeper insights into the mechanisms underlying the evaluation of and response to affective content, resembling the algorithm-driven content delivery experienced by users on real-world social media platforms.

Another limitation pertains to the assessment and range of measured consequences. In the conducted experiments, the focus was solely on participants' intention to share content on social media. However, it is important to acknowledge that there may be disparities between intention and actual sharing behaviour among real social media users. Unlike participants in controlled experiments, actual users do not necessarily face reputational consequences or gain any benefits from sharing content. Therefore, future research should consider investigating actual sharing behaviour to obtain a more comprehensive understanding of the dynamics at

play. Furthermore, it is crucial to recognize that the range of news articles provided in the experiments was confined to a specific style, predominantly local news. It is plausible that the sharing patterns for this particular style may differ from those observed for political news, sports events, or private news. Based on the findings of Study 1 – 3, negative political news might be shared more than positive ones, while the opposite trend may be observed for sports news (Gruzd et al., 2011) or scientific articles (Milkman & Berger, 2014). Thus, it is important for future studies to incorporate a broader range of content types to capture the nuanced variations in sharing behaviours across different news categories. To address both of these concerns, future research should aim to capture a more realistic version of content and user behaviour by measuring social media users' actual emotion perception while they engage with their real social media accounts. This approach would provide a more authentic representation of the content users encounter and the context in which they interact with it.

Beyond the behavioural consequences which were the focus of these experiments, future work could also include measures of subjective well-being. While existing literature suggests that exposure to positive content can enhance well-being (Keltner, Bowman, & Richards, 2017), and exposure to negative content can diminish it (Mannell & Meese, 2022; Potts & Sanchez, 1994), this study did not directly test these potential consequences of exposure. The rationale behind this omission was that each trial in the experiment could have been either positive or negative, and by mixing these trials, the positive effect of the positive trial might have masked the negative effects of the negative trials.

Regarding the limitations and implications of the findings in relation to the ensemble coding literature, I propose that people tend to overestimate the intensity of emotional language sequences because they perceive more emotionally intense responses as more representative, primarily due to their heightened attention to emotional words. Although the

experiments conducted in this chapter did not directly measure participants' attention to emotional words, they provided indirect evidence of this attentional bias through the word-only study (Study 9, see Appendix of Chapter 5). Future studies could further explore this by incorporating eye-tracking methodologies to empirically demonstrate the attentional bias towards emotional words.

Finally, having successfully presented evidence demonstrating overestimation in complex judgments relating to emotional language in sentences, the logical progression is to expand these findings to encompass other elevated forms of judgment, including trust and morality. To do so, researchers can employ a similar paradigm to examine these overestimations by, for instance, presenting participants with a series of news articles concerning an institution or an individual, and subsequently gauging their perceptions of the moral or trustworthy nature of the subject based on the sequence of information they have received.

To conclude, this chapter's research establishes two important observations that are significant in the context of this thesis. First, this research demonstrates the significant consequences associated with consuming affective content on social media. The act of engaging with such content profoundly influences how we assess the emotional intensity of events, as well as people's opinions regarding them. Unfortunately, the sequential presentation of this content appears to amplify both the perceived intensity of the events we encounter and our judgments of others' expressions. Second, I verify that the intensity of emotional language indeed increases the likelihood of content being shared, although there was no evidence suggesting that people engage more with negative content.

CHAPTER 6. General Discussion

General overview of the chapter

In this thesis, I have presented a series of eight studies intended to answer two main questions. The first question focused on determining the most prevalent type of affective content on social media by examining the content that is most frequently created and shared. Chapters 2-3 focused on analysing the prevalence differences between negative and positive affective content, with a particular focus on the potential harm to users caused by negative affective content. The second question focused on the consequences of affective content on social media. Chapters 4 and 5 provided insights into these consequences from the perspectives of both the content sender (Chapter 4) and the content perceiver (Chapter 5). The primary objective of these chapters was to gain a deeper understanding of the effects experienced by individuals engaged in creating and consuming affective content on social media.

Throughout the four empirical chapters, I aimed to create a more comprehensive understanding of the role and consequences of affective content on social media by illuminating various perspectives, including those of content sender and their perceivers, within various contexts. In pursuit of this goal, I employed a combination of computational and experimental designs, capitalizing on the strengths inherent in each approach. Below, I present the key findings related to each of the two main research questions addressed in this thesis in turn and for each of them separately provide an overview of the methodological contributions made by the empirical work. For each question, I then discuss the implications

and limitations of my findings, before identifying the most promising avenues for future research. I conclude by discussing broader theoretical considerations.

Prevalence of Positive and Negative Affective Content on Social Media

The first objective of this thesis was to investigate the prevalence of negative and positive affective content on social media and understand which type is produced and shared more frequently. Based on prior literature, I hypothesized that although positive affective content is created more often, it is negative affective content that tends to be shared more (Bellovary et al., 2021; Fine & Hunt, 2021; Tsugawa & Ohsaki, 2015). While there are instances where positivity is shared more, such as in scientific articles (Milkman & Berger, 2014), Olympic games posts (Gruzd et al., 2011), or news articles distributed via email (Kraft et al., 2020), my general hypothesis was that negative affective content would be more commonly shared across most contexts and by most users due to a supposedly universal negativity bias (Bebbington, MacLeod, Ellison, Fay, & Behavior, 2017). To test this hypothesis in different scenarios, each study in my thesis either focused on a specific context or examined the role of different user types. Previous studies have predominantly centred around negatively perceived or contentious political events and tweets posted by public figures, which may have contributed to the general conclusion that negativity is more extensively shared than positivity. In my studies, I aimed to understand the boundary conditions surrounding the over-sharing of negativity by exploring both positive and negative contexts, political and non-political topics, as well as diverse user types such as public figures and ordinary users.

Findings

Study 1 Sharing of Affective Content Produced Public Figures Compared to Ordinary Users. In Study 1, I focused on the role of user types by comparing public figures

with ordinary users to understand which type of affective content is produced by them and shared by other users. Additionally, I analysed the user characteristics that influenced the likelihood of their negative affective content being shared. The findings revealed that public figures, despite primarily creating positive affective content, were more likely to have their negative affective content shared. This stronger association between negativity and sharing was consistently observed across all types of public figures including news media outlets, politicians, and journalists. However, I did not find evidence to support the notion that negativity was shared more than positivity when produced by ordinary users. Moreover, two specific user characteristics, the number of followers and the proportion of political content, fully mediated the relationship between user type and the likelihood of their negative affective content being shared. Notably, the number of followers emerged as the more influential factor in this mediation model, as it exhibited greater variation between the two user types and showed stronger mediation in the overall model. These results supported the notion that most users, particularly public figures, in fact, create more positive affective content. Contrary to the assumption that negative affective content is generally shared more on social media, this study provided evidence that such prevalence is limited to public figures or users with a high number of followers who frequently engage in political discussions. For other users, both high-intensity positive and negative affective content was shared to a similar extent.

Study 2 – 3 Content Sharing by Users with Positive or Negative Evaluations of a Situation. To address the research gap regarding the sharing of negative affective content in political contexts, I conducted Studies 2 and 3 to compare the creation and sharing of affective content concerning positive and negative events. The aim was to determine if previous research on political contexts had overly focused on negative events and public

figures or whether negative affective content is consistently shared more generally in political contexts. In Study 2 and Study 3, I examined the sharing patterns of affective content in predominantly negative and predominantly positive situations. Specifically, Study 2 looked at a situation (presidential election) from the perspective of both the winning side and the losing side, and Study 3 looked at two separate situations that were predominantly positive or negative for users who posted about them on Twitter. The findings from both studies indicated that negative affective content tended to be shared more, regardless of whether the situation was negative or positive. To challenge the argument that negative affective content shared during political celebrations primarily originated from outgroup members, I conducted an analysis of political affiliation in both studies. The results indicated that the users generating the negative affective content during political celebrations were primarily ingroup members (conservatives in Study 2 and liberals in Study 3). Additionally, an analysis of the content of the negative affective content written in response to the celebrations revealed that negative language was primarily used to describe hardships, and past obstacles, or to portray the outgroup in a negative manner. Taken together, the findings from these two studies demonstrate that even for ordinary users, negative affective content is shared more frequently than positive affective content following political events, irrespective of users' support for or opposition to the outcomes of these events. This suggests that the sharing of negative affective content in political contexts is not solely driven by certain user types such as public figures as was the case in Study 1, but rather reflects a broader trend in how affective content is disseminated in political discussions online generally.

Study 4 Affective Social Sharing During an Ongoing Crisis. To investigate whether negative affective content is more prevalent in non-political events that are appraised as negative by most people, I conducted an analysis of tweets pertaining to the ongoing

COVID-19 pandemic. The results of the study revealed that following the onset of the pandemic, there was an increase in the use of negative language, accompanied by a decrease in the usage of positive language. These findings underscore the inherent negativity associated with this global crisis. However, in spite of the increasing prevalence of negativity, it is noteworthy that users continued to produce and share a greater amount of positive than negative affective content. Interestingly, the negative affective content that got shared received a higher total number of retweets than the positive affective content. This finding again challenges the hypothesis formulated in Chapter 1, which was based on prior literature and presumed that negative affective content is predominantly shared on social media. Instead, the results reveal a significantly more positive trend regarding the content that is both produced and shared.

Study 5 –8 Evaluation of Emotion Expressions in Text Sequences. My hypothesis concerning the sharing of affective content was built upon two key assumptions. First, content that employs more intense emotional language captures greater attention, thus resulting in more sharing. Second, due to the negativity bias, negative affective content receives more attention than positive affective content. As a result, I anticipated the finding that negative affective content is shared more than positive affective content. Studies 5-8 were designed to evaluate whether the assumptions underlying my hypothesis were in fact correct. Specifically, these four studies aimed to determine whether content perceived as more emotionally intense is shared more frequently and whether negative affective content is shared more than positive content. The results revealed that while the *intensity* of emotional language was indeed associated with a higher inclination to share, no significant difference was found between positive and negative *valence*. These results are in line with those of the

other studies of this thesis, indicating that while intensity plays a consistently crucial role in content sharing, negativity alone does not suffice.

Implications of the Findings Relating to the First Main Research Question

The findings of this thesis regarding the prevalence of created and shared affective content on social media emphasize the important effects of context and content producers on the prevalence of affective content online. They have implications for social media platforms, their users, and researchers investigating affective content on social media. These implications encompass (1) the creation of content, (2) the sharing of content, and (3) the underlying mechanisms that drive the prevalence of affective content on social media.

In relation to content creation, findings from Studies 1 - 4 present a favourable outlook on social media platforms and their users, by indicating that original content tends to be more positive compared to shared content. Moreover, even amidst ongoing crises as observed in Study 4, social media users displayed a remarkable inclination towards positivity, which supports the concept of a positivity bias in expression (Dodds et al., 2015). This increased positivity is particularly prominent for public figures or users with a substantial number of followers, aligning with the notion of reputation management (Citra, 2018; Lin et al., 2014). If social media users desire to enhance their exposure to positive affective content, one feasible piece of advice for them is to primarily view original rather than shared content and follow users with the higher number of followers who tend not to produce political content.

Moreover, consistent with prior research, my findings suggest that social media expressions do partially track the public mood (Eichstaedt et al., 2018; Jaidka et al., 2020; Otto & Eichstaedt, 2018). Specifically, events or situations that are generally perceived

positively by most individuals, such as political victories, lead to a decrease in negative affective content and an increase in positive affective content. Conversely, events such as the COVID-19 pandemic or political defeats lead to the opposite trend, although it should be noted that negativity does not consistently exceed positivity to a significant degree, as mentioned earlier. These findings indicate that when using social media as an approximation of public mood, it is necessary to compare the affective tone of content to a baseline rather than solely relying on the absolute difference between positive and negative content.

Shared social media content contains a relatively higher proportion of intense affective content than unshared content, but only contains a higher proportion of negative affective content when produced by certain users in certain scenarios. Positive affective content produced by ordinary users outside of politics is shared as much as their negative affective content. On the other hand, negative affective content produced by users with a large number of followers (such as public figures or in political contexts) is shared more often than their positive affective content. The resulting inflation of negative affective content could have harmful consequences for individuals (Nguyen, Glück, & Jackson, 2022) and collectives (de Mello et al., 2022). A user's exposure to negative affective content depends not only on the users they follow (direct connections) but also on who the person they follow is following (indirect connections). This implies that even if a user is not directly connected to users with large numbers of followers, such as public figures or users who frequently engage in political discussions, they may still encounter shared negative affective content via indirect connections. As detailed in Chapter 2, direct connections of public and political figures retweet and disseminate a significant amount of their negative affective content. To safeguard users against potentially undesirable negative content from individuals they do not directly follow, social media platforms could introduce an optional feature. This feature

should enable users to exclusively view content from users they actively follow, empowering them with greater control over their content consumption.

Why do people share and create affective content on social media? Examining these patterns of creating and sharing affective content provides insights into the mechanisms underlying such behaviour. It appears that users' inclination to create more positive affective content is influenced by various factors, including the topic addressed, the context in which they are sharing, their need for impression management (Luo & Hancock, 2020), and the level of their positivity bias (Dodds et al., 2015). Additionally, in the following section, I explore interpersonal emotion regulation as a potential motivation for producing affective content in times of collective crisis.

The mechanism underlying the sharing of affective content appears to be significantly more intricate than previously believed. In Chapter 1, I put forward the argument that one of the factors driving content sharing is the higher level of attention that is directed at more intense affective content, which subsequently leads to increased sharing. Moreover, I hypothesized that negative affective content is shared more frequently than positive affective content as a result of the negativity bias, which is the inherent inclination to pay more attention to negative stimuli (Lazarus, 2021). While I discovered evidence supporting the idea that a higher intensity of emotional language in content is linked to the intention to share, I did not find any difference in sharing intent between positive and negative valence. Therefore, my findings do not support the notion of a universal negativity bias when it comes to sharing. Nevertheless, my findings about the network features and contexts in which negativity sharing is more prevalent provide clues about what the mechanisms might be in those contexts. Rather than assuming a universal inclination towards negativity, two other potential explanations for the frequency of sharing negative affective content emerge. First,

negativity sharing seems to be moderated by tie strength (Hansen et al., 2011), with negativity shared more between weaker ties, while positivity is shared more between close ties (Berger & Milkman, 2010; Fan et al., 2016). Second, the sharing of negative affective content is context-specific. In particular, negativity is disproportionately shared in political contexts, whereas this pattern does not hold true in non-political contexts, regardless of individuals' appraisal of the context. In political contexts, users are often motivated by intergroup hostility and considerations of reputation, potentially leading them to share more negative content (see outgroup animosity, Rathje et al., 2021). It remains unclear whether there is a common underlying cause for the increased sharing of negative affective content between weak-ties and in specific contexts such as politics, which I discuss below in the section on *Broader Theoretical Considerations and Future Directions*. In sum, these findings suggest that it is important to employ a more nuanced perspective when investigating the prevalence of affective content on social media.

Limitations and Future Directions

Despite my effort to comprehensively understand the prevalence of affective content on social media by employing a combination of experimental and observational approaches, there exist certain limitations of the research I have presented on this topic and unanswered questions that should be addressed in future research.

One limitation is that even after quantifying the affective content that is produced and shared, I was still unable to definitively determine the amount of affective content that users are exposed to as this depends on factors such as the exact times when they are online and how this content is curated by Twitter's algorithm which determines what content is displayed to users. Social media users report that they do believe that social media especially

facilitates the spread of harmful content such as misinformation, hatred, and content containing intergroup conflict (Rathje et al., 2022). This lay belief is supported by the argument of functional misalignment, which suggests that algorithms tend to amplify existing tendencies, many of which lean towards negativity and intergroup conflict (Brady et al., 2023; Milli, Carroll, Pandey, Wang, & Dragan, 2023). Conversely, contrasting findings propose that the algorithm's role in political polarisation may be less significant than that of users' pre-existing tendencies and consequent actions (Feezell et al., 2021). To more accurately assess engagement and consumption of negative content, future research could consider utilizing actual consumption data, such as analysing continuous screenshots of individuals' phone usage (for more detail, see Brinberg et al., 2021).

The reliance on the simplistic categorization of affective content into positive and negative types using sentiment analysis gives rise to two limitations. First, it raises questions about the extent to which affect expression accurately reflects subjective feeling. This topic will be addressed in a later section when discussing the choice of sentiment analysis tools. Second, the binary categorization itself poses a limitation. Specifically, expressions can encompass various discrete affective categories, including anger, sadness, happiness, contempt, pride, and more. However, the present research consolidates these categories and concentrates solely on valence and intensity. Previous studies have highlighted that different types of affective content are shared to varying degrees (Fan et al., 2016; Mercadante, Tracy, & Götz, 2023). Even within the category of negative affective content, there are differences in the extent to which specific affect types spread. For example, types of high arousal affect such as fear (Saha et al., 2023), incivility (Frimer et al., 2023), and anger spread further than low-arousal types such as sadness (Fan et al., 2016). When examining the prevalence and the consequences (as discussed in a later section) of sharing negative content, it could

consequently make a significant difference whether anger or sadness is more prominently expressed and shared. Therefore, future research could distinguish more specific types of positive and negative affective content to gain a more comprehensive understanding of affect sharing on social media.

Finally, the research presented in this thesis also does not directly assess the motivations behind sharing and what function sharing serves in the eye of the sharer. Based on previous research, I assumed that sharing without editing or commenting generally indicates agreement with the content as the sharing user just passes on the information in its original form (Goldenberg & J. J. Gross, 2020). However, this research does not explore or distinguish the exact motivations behind sharing, which could include supporting the original creator, raising awareness, mobilizing people for action, or simply signalling concern for a particular topic. One example that highlights how sharing may not accurately reflect users' actual opinions or exact agreement as assumed in this thesis is within political contexts. As mentioned earlier, group identity is salient in the political context, and users might be motivated to signal their group affiliation. To make their ingroup membership evident to others, individuals may share content that they believe reflects their ingroup's opinion. However, peoples often overestimate the strength of opinions within their ingroup (Ahler & Sood, 2018; Goldenberg et al., 2023; Robinson, Keltner, Ward, & Ross, 1995), leading them to adopt more extreme positions than their true opinion when sharing content. Understanding different motivations could shed light on the mechanisms that drive users to share certain affective content over others. To explore this further, future research could employ experimental manipulations that activate specific motivations, such as making group identity salient, to induce the motivation of signalling ingroup membership. Observing what participants share when these motivations are active, it would allow for an evaluation of

whether distinct motivations for content creation or sharing are associated with specific types of affective content.

Overall, the findings of this thesis contribute to a more nuanced understanding of the factors that influence the prevalence of specific types of affective content. Unlike previous research that primarily focused on analysing isolated properties of individual tweets such as the use of moral language in one specific context, this thesis aimed to compare the prevalence of affective content across different contexts and user types. The results underscore the complexity of the sharing pattern and the importance of context and producer, ultimately suggesting moving beyond the universal assumption that negativity is consistently shared more frequently. The main limitations pertain to generalizability as the research of this thesis was centred around the comparison between the binary categories of positive and negative affective content, and lacked an in-depth examination of the motivations driving content sharing. The limitations arising from the use of dictionary-based tools, exclusive reliance on Twitter data, and the sole focus on text data without considering image data, are further discussed in a later section.

Despite these limitations, this work provides valuable insights into when negative affective content can be expected on social media platforms. This knowledge can assist social media platforms and users in developing strategies to potentially reduce the frequency of negative affective content. Broader issues regarding the implications and desirability of negative affective content, as well as affective content in general, are explored in the next section.

Consequences of Affective Content for Sender and Perceiver.

The second objective of this thesis, following the quantification of the prevalence of affective content on social media, was to assess the potential consequences associated with such content. Recognizing that creating and sharing content may have different consequences compared to consuming content, I further divided this question into the consequences for the sender (Chapter 4) and the perceiver (Chapter 5). To test the consequences for the sender in Study 4, I adopted a social sharing perspective that emphasized the functional reasons behind expressing affective content to others (Rimé et al., 2020). Despite the potential benefits of SSE, previous research also suggested potential harmful consequences of SSE (Rimé et al., 2010). In contrast, when examining the consequences of consuming affective content, I did not assume that it is beneficial to engage with such content, especially when it comes to negative affective content. Drawing on emotion contagion theory (Hatfield, Hatfield, Cacioppo, & Rapson, 1994), it is posited that consuming affective content could lead to convergence between the perceived affect expressed in the viewed content and the viewers' own emotional state. Consequently, I sought to understand in Study 5 – 8 how content on social media is perceived in order to gain insight into the potential consequences of consuming affective content. Given that affective content is typically presented in a sequential feed-like format, the primary objective in understanding the consequences for the perceiver was to investigate how sequences of affective content are perceived by viewers.

Findings

Study 4 Affective Social Sharing During an Ongoing Crisis. The primary objective of Study 4 was to examine the consequences of expressing and sharing affective content on social media in the context of the COVID-19 pandemic. This particular context enabled me to

address two research gaps pertaining to (1) exploring both the beneficial and harmful outcomes of social sharing, and (2) examining its effects during ongoing events. Previous research has primarily focused on acute negative events and the impact of social sharing on people's emotional states following the completion of such events. Moreover, the global nature of the pandemic allowed for an examination of the consequences of social sharing in events that were not localized, such as the French or Norwegian terror attacks (M. Eriksson, 2016; Garcia & Rimé, 2019). The findings indicated that sharing more pandemic-related content (SSE) appeared to contribute to Twitter users feeling less negative and more positive about the pandemic, aligning with the emotional recovery hypothesis. Additionally, temporal causality analysis provided support for the claim that social sharing preceded an improvement in expressed affect among Twitter users. Overall, this study provides evidence suggesting that expressing affective content may serve a functional role and aid in emotional recovery for content producers.

Study 5 – 8 Evaluation of Emotion Expressions in Text Sequences. The main objective of these experiments was to investigate how individuals perceive affective content online when it is presented to them sequentially. Building upon existing research on the perception of emotions expressed in faces (Goldenberg, Schöne, et al., 2022; Goldenberg, Weisz, et al., 2020), I hypothesized that people would tend to overestimate the emotional intensity of texts. This assumption was based on the idea that individuals pay more attention to and therefore assign greater weight to emotionally intense stimuli when making judgments across multiple stimuli. Confirming this hypothesis, my findings revealed that participants did indeed overestimate the average emotional intensity of the language expressed in sequences of text responses to fictional news articles. Moreover, this overestimation was more pronounced for longer sequences. Analysis of how the context of text responses

influenced the ratings of both the context and the sequence revealed a reciprocal influence between the two. This reciprocal influence could enhance the perceived emotional intensity of the language in these estimations. As mentioned in the previous section, individuals demonstrated a stronger inclination to share affective content that they perceived to express stronger emotions, providing evidence that the perception of affective content has behavioural consequences. When exploring potential explanations for this effect, I discovered evidence suggesting that participants did not exhibit a systematic memory bias towards responses that are more emotionally intense. However, they may have weighted the more intense responses more heavily in their ratings because they perceived them as more representative or diagnostic of the overall sequence. In summary, these experiments revealed three noteworthy consequences. First, due to the sequential presentation of stimuli, individuals tend to overestimate the intensity of affective content. Secondly, this estimation influences the perception of the content, and vice versa, creating a reciprocal relationship. Lastly, affective content that is judged as more intense is more likely to be shared regardless of its valence.

Implications of the Findings

Ongoing debates surrounding the benefits and harms of mainstream social media usage seem to indicate that consequences depend on how individuals engage with social media and the content they choose to interact with (for a review, see Kross et al., 2020). A commonly recognized differentiation in kinds of social media use is between passive observation and active content creation and interaction, with the latter being found to be more beneficial to users (Liu, Baumeister, Yang, & Hu, 2019). Consistent with these findings, Study 4's results suggest that actively creating and sharing content, even in response to

negative events, is associated with emotional recovery. Unlike offline face-to-face conversations, social media provides users with easily accessible audiences at all times, potentially offering emotional support and fostering a sense of connectedness with others (de Mello et al., 2022). Positive psychology research has identified prosocial behaviour and feelings of connectedness as contributors to improved well-being (Nelson et al., 2015; Regan, Radošić, & Lyubomirsky, 2022). It, therefore, seems plausible that individuals can utilize social media to enhance their social interactions and generate positive outcomes through the online expression of both positive and negative affect.

Political discourse might play a distinct role in affecting the well-being of social media users, as it carries various risks that differ from other forms of online content. The current political climate, particularly in the United States, has become highly polarized (Brady et al., 2017; Finkel et al., 2020), and characterized by animosity towards outgroups (Rathje et al., 2021). Consequently, users not only generate more negative affective content but also their negativity is further amplified through disproportionate sharing following political victories as well as in political defeats (as shown in Studies 2 – 3). Additionally, the reinforcing nature of social media, facilitated by social learning (Brady et al., 2021; Lindström et al., 2021) and algorithms, that encourage user engagement (Brady et al., 2023), can perpetuate these trends, potentially explaining the observed escalation of incivility in political contexts over time (Frimer et al., 2023). As a result, dissemination of affective content in a political context might be especially harmful to individuals' well-being as well as intergroup relations. For example, an experimental study has shown that engagement with political content is associated with a decline in well-being, whereas refraining from such content can have a positive impact on well-being (Feinberg et al., 2020).

Instead of condemning the use of social media for political discourse solely based on the associated risks to well-being, it is imperative to carefully consider the role of negative affective content in such discussions. On one hand, social media platforms often foster the formation of echo chambers, where individuals tend to congregate and primarily interact with like-minded individuals (Boutyline & Willer, 2017; Petrov & Proncheva, 2018; Wojcieszak, Casas, Yu, Nagler, & Tucker, 2021), which can contribute to the growing political polarization by restricting exposure to diverse perspectives and exacerbate the negative outcomes of political discourse (Bossetta, 2018; E. Peterson, Goel, & Iyengar, 2018). Importantly, negative political expressions online can translate into offline behaviours. For instance, a study discovered a correlation between anti-refugee sentiment on Facebook and crimes against refugees in municipalities with higher social media usage, despite their similarity to municipalities with lower social media usage along other relevant dimensions such as access to the internet and average education level (Müller & Schwarz, 2021).

On the other hand, it is important to acknowledge that social media provides a platform for reporting important negative events at a faster pace than traditional news outlets (Fitzpatrick, 2018). Ordinary people can share real-time updates, making social media particularly crucial for political movements like #MeToo, BlackLivesMatter, and the Arab Spring, where it has played a vital role in driving positive societal change (Manikonda, Beigi, Kambhampati, & Liu, 2018; Mundt, Ross, & Burnett, 2018; Wolfsfeld, Segev, & Sheafer, 2013). Furthermore, negativity plays a crucial role in political engagement, particularly hatred (Halperin, Canetti, & Kimhi, 2012), can serve as a catalyst for mobilizing individuals (Fischer, Halperin, Canetti, & Jasini, 2018). This is exemplified by the prevalent use of negative campaigning, prominently spearheaded by Donald Trump on social media during the 2019 general election (J. H. Gross & Johnson, 2016). Negative campaigning is a political

strategy in which candidates or political parties direct their efforts toward criticizing opponents instead of solely highlighting their own strengths and policies. In summary, when it comes to negative affective content in the political context, there exists a trade-off between its functional role in garnering support (Federico, Johnston, & Lavine, 2014), and the potential risks it poses in relation to well-being and polarization.

In addition to considering our level of exposure to positive and negative affective content, it is important to recognize that the manner in which we consume such content makes a difference to the consequences of social media use. My research reveals that users tend to amplify the perceived intensity of affective news articles and their corresponding text responses when they are presented sequentially, as is often the case in social media feeds. Consequently, the information we receive through social media may appear more intense than if we had encountered each item separately. This distorted perspective can significantly skew our understanding of events and of others' appraisal of them, regardless of whether the events are negative or positive. Without further evidence, it would be premature to conclude that this heightened perception of intensity has detrimental effects on social media users in all contexts. Evidence indicating that individuals with higher emotional intelligence, characterized by their ability to accurately assess emotions, experience improved physical and mental well-being (Martins, Ramalho, & Morin, 2010), may indirectly imply that misperceiving emotional intensity could be maladaptive. While the exact consequences require further investigation in future research, previous research as well as Studies 5 – 7 provide evidence suggesting that social media usage can profoundly influence our perception of the world (Brady et al., 2022; Goldenberg, Abruzzo, et al., 2022; Goldenberg et al., 2023).

Given the prevalence of positive and negative affective content and the tendency to overestimate its intensity, it is crucial to critically review the assumption that it is desirable

for affective content to become less negative and more positive. This assumption overlooks the possibility of positive consequences of negative affective content consumption on social media. For instance, exposure to content conveying someone else's negative experience has the potential to elicit a positive emotional reaction in individuals such as a sense of similarity or empathy (Luminet IV, Bouts, Delie, Manstead, & Rimé, 2000). Sharing negative experiences with others can also facilitate social bonding (Newson, Buhrmester, & Whitehouse, 2023). Conversely, exposure to positive affective content showcasing "happy others" can trigger negative emotional states, such as feelings of depression or a sense of inferiority because of the unfavourable social comparison (Alfasi, 2019; Parsons, Alden, & Biesanz, 2021; Uhlir, 2016). Thus, emotional reactions to affective content can be influenced by the interaction between the valence of the content and individual differences for example social comparison tendencies of the message receiver (Park & Baek, 2018). Acknowledging these complexities helps us to avoid overlooking the potentially positive effects of consuming negativity and the negative effects of consuming positivity, and it becomes evident that claims about the desirability of positive or negative affective content do not apply universally.

Limitations and Future Directions

While the research reported in this thesis sheds light on the impact of affective content on both content sender and perceiver, it is essential to acknowledge the existing limitations and unanswered questions that necessitate further investigation in these domains. One primary limitation is the limited assessment of subjective well-being in this thesis. In Study 4, the affective consequences of SSE were only indirectly measured by examining changes in users' expressions of affect. Although previous research has used similar measures as proxy

indices of actual affective states (Bollen et al., 2009; Otto & Eichstaedt, 2018), and despite the similar responses observed in the survey study where affect was more directly self-reported, there remains a need for more accurate and direct assessment of well-being. Well-being encompasses at least three components: positive emotions, negative emotions, and life satisfaction (Diener et al., 2017), but the research reported in this thesis has only indirectly assessed the emotion facets. Further, Studies 5 - 7, which explored the consequences of perceiving affective content, included no assessment of any components of well-being. Future studies investigating the consequences of sharing and viewing affective content should address this limitation by assessing well-being using more direct measures such as the Subjective Well-being Scale (Diener et al., 2010) and the Warwick-Edinburgh Well-being Scale (Stewart-Brown & Janmohamed, 2008).

The consequences of creating, sharing, and viewing affective content within the realm of online politics warrant further attention, as they have implications not only for individual well-being but also for political discourse and democracy more generally. It is essential to conduct additional research that explores how to sustain people's interest and engagement with politics without exacerbating political polarization and intergroup conflicts. Although the research presented in this thesis did not specifically analyse the consequences of political social sharing or the viewing of political content, future research could employ similar approaches to examine these aspects in political contexts. This would provide a better understanding of how social media can be utilized to facilitate political discourse without deepening the divide between political partisans.

Finally, I did not further examine the role of other factors that could potentially influence the consequences of creating, sharing, and viewing affective content. For example, the age of social media users appears to have an impact on their well-being, with stronger

effects for younger adolescents (Orben, Przybylski, Blakemore, & Kievit, 2022). Additionally, there are individual differences such as the tendency for social comparison (Luo & Hancock, 2020), other personality traits (Bunz, 2021; Gerson, Plagnol, & Corr, 2016), and cultural variations (LaRose, Connolly, Lee, Li, & Hales, 2014) that are associated with differences in how individuals are affected by social media use. Therefore, in future studies examining the creation and consumption of affective content, it would be valuable to explore how these developmental, individual, and cultural differences might moderate effects on different outcomes.

In summary, my research on affective content online has highlighted a useful distinction between consequences for the sender and the perceiver of such content. The findings of my research align with previous studies, emphasizing that the benefits or harms arising from social media use are contingent upon individuals' engagement with the platform. Regarding the sender, expressing affective content seems to be associated with emotional recovery, for example during times of collective crisis. This suggests that actively utilizing social media platforms is potentially beneficial for its users, as further evidenced by the positive correlation between greater engagement in social interactions and an increased sense of belonging, and improved well-being (Bowden-Green, Hinds, & Joinson, 2020). In contrast, for the perceiver, frequent exposure to sequential affective content on social media is found to lead to a misperception of the affective intensity of online content. Passive consumption of content, without active participation in content creation, has been associated with decreased well-being and a heightened sense of moral outrage (Brady et al., 2022; de Mello et al., 2022). The main limitations of the present research that are discussed concern the limited assessment of consequences such as well-being, and the lack of further examination of specific topics, including politics. Finally, future studies should look at the

individual differences in social media behaviour and how they affect users' vulnerability to negative consequences. It is crucial to note that social media use can be harnessed to improve users' lives, and subsequent research should concentrate on discerning methods to realize this positive potential.

Assessment of Affective Content and Engagement

In Chapter 1, I explored some of the methods and challenges that research on analysing affective content and people's engagement with it entails. I also laid out the rationale behind my selection of specific analysis tools and measures of engagement. While comparing different methods was not the primary focus of this thesis, I nevertheless derived valuable insights about the methodology while addressing the research questions. In particular, in this section, I discuss the findings related to the assessment of affective content, the measurement of engagement, and the modelling of observational data derived from social media.

Findings and Implications

Assessment of affective content. In this thesis, I defined affective content as such content that uses valuation words and affective expressions. To evaluate affective content, researchers have the option to employ various analysis tools, including dictionary-based (Hardeniya & Borikar, 2016), supervised (Samal, Behera, & Panda, 2017), and unsupervised machine learning tools (Singh, Piryani, Uddin, & Waila, 2013). However, it is not always clear which specific linguistic features machine-learning methods use to identify relevant content, so I made the decision to primarily utilize dictionary-based tools. This choice

ensured that, in accordance with my definition, affect estimation was solely based on affective words while excluding the influence of other linguistic features. As I employed the unsupervised machine learning technique of topic modelling to gain a deeper understanding of the content of tweets in Studies 1-3, I was able to illustrate some of the advantages and disadvantages associated with the decision to choose dictionary-based approaches.

In Study 1, I found justification for my decision to rely on a dictionary-based approach to assess affective content when I employed topic modelling to identify political tweets. While most of the words within the political topic were accurately captured and appeared to carry political meaning upon manual inspection (see Appendix of Chapter 2, Table 2.5), some other words such as "time," "day," and "feel" were also identified as highly associated with the political topic, despite having no obvious political connotations. As a result, there were also sentences classified as political that would not have been identified as such by human annotators (see Appendix of Chapter 2, Table 2.6). On the other hand, I also discovered the disadvantages of dictionary-based approaches while analysing tweets classified as negative in political celebrations surrounding same-sex marriage in Study 3. The words "cry" and "tear" were frequently retweeted in tweets classified by the dictionary-based approach as negative, despite expressing positive affect in this context (tears of joy and crying out of happiness). Fortunately, even after removing these results, the overall findings remained consistent. Nevertheless, this highlights a limitation of dictionary-based approaches in capturing complex language and expressions beyond their direct meaning.

Even within the chosen approach of using dictionary-based tools, it is important to note that not all tools are equally suitable for all specific research questions and analysis plans. Throughout Studies 1-4, I conducted the analysis using at least two different analysis tools, primarily VADER and SentiStrength. Encouragingly, in most cases, the results from

these tools were consistent. However, there were still discernible patterns where one tool outperformed the other. One of the key differences to consider in this regard is how the positive and negative scores are calculated. As discussed in Study 1, VADER calculates a total sentiment score ranging from -1 (very negative) to +1 (very positive), along with separate positive, negative, and neutral scores that range from 0 to 1 in intensity. These three sub-scales are interdependent and always sum up to 1. In contrast, SentiStrength does not provide a total sentiment score, and its positive and negative sub-scales operate independently. The consequence is that in cases in which I was interested in overall sentiment and its relationship with sharing, VADER performed better, as shown in Studies 1 and 4. Conversely, SentiStrength demonstrated superior performance over VADER when it came to calculating separate positive and negative language scores for each individual tweet (Study 2-3), with the exception of Study 4 (see Chapter 4 for more details). Hence, it is crucial to judiciously choose to select the appropriate analysis tool based on the specific research question and the nature of the data being analysed, factoring in both statistical fit parameters and theoretical considerations.

Online engagement with affective content. The metric of engagement that I used in this thesis was the number of retweets, for two primary reasons. First, retweets can be interpreted as an indication of agreement with the content's message, as users choose to pass it on without making any further modifications (Goldenberg & J. J. Gross, 2020), although the extent to this agreement might vary depending on peoples underlying motivation (see *Limitations in the Section Prevalence of Positive and Negative Affective Content on Social Media*). On the contrary, the act of "liking" introduces greater ambiguity. Likes are only immediately visible to the content creator, making this act less public and thus reducing repercussions for openly endorsing a certain message. As a result, users might not even read

the content, but merely express support for the content creator by liking the post. Second, by retweeting content, users contribute to its prevalence by sharing it with their own followers, while this is not the case for likes. The number of likes is confounded with retweets because if a tweet is retweeted frequently, more people will see it and thereby get the chance to "like" it. This confounding factor can potentially lead to misleading results and interpretations. To illustrate, let's assume that each item of positive affective content has a greater probability of receiving a like than each item of negative affective content. However, let's also assume that negative affective content is more likely to be retweeted, meaning that the overall number of likes for negative affective content is higher than it would otherwise be. Under these circumstances, we might find similar frequencies of "likes" for positive and negative affective content, even though positive affective content is actually more liked. In Study 2-3, I also analysed the relationship between likes and the use of emotional language (Appendix of Chapter 2). However, the pattern of results appeared to be unsystematic between the examined situations, which may be attributed to the aforementioned confounding effect of retweets. Alternatively, Frimer et al. (2023) proposed using the like-to-retweet ratio as a potential alternative. While this approach addresses the confounding influence of retweets on likes, it reintroduces concerns over the interpretation of "likes". Given the uncertainty surrounding whether users agree with or even read the content they've liked, I decided not to include this measure in this thesis. Additionally, the first question of my thesis was to quantify the prevalence of content, which is exclusively influenced by retweets rather than likes. Therefore, without further research into the meaning and implications of likes, using the number of retweets as a metric of engagement remains conceptually clearer than attempting to use the ratio between them.

The association between emotional language intensity and engagement had a generally low effect size across studies. For instance, Studies 2-3 found approximately 1.8% to 4.8% increases in retweets per unit increase in the intensity of negative language. This finding aligns with previous research that has reported effects ranging from 2% to 6% (Brady et al., 2017; Rathje et al., 2021). On the other hand, positive language intensity was only reliably associated with sharing in Study 3, with an increase in retweets of 1.3% per unit increase in intensity. As discussed in Chapter 3, although these effects may appear small, they hold practical significance considering the vast number of social media users and the overall volume of social media activity. Additionally, the findings reported in Chapter 2 showed that influential users such as public figures and individuals with a large number of followers exhibited stronger associations between language intensity and sharing, as well as a higher overall sharing frequency due to their follower base, making it evident how small effects can scale up with increased social media activity.

Modelling of observational data derived from social media. Using observational data, particularly language data, to model behaviour poses numerous challenges. Some of these challenges stem from the unique characteristics of social media data, while others arise from the specificities of natural language. One outstanding challenge related to natural language relates to the large disparities in the commonly found distribution of words, which can prove especially difficult for dictionary-based approaches. The frequency of words in natural language follows Zipf's law, which states that the frequency of any given word in a language is inversely proportional to its rank in the frequency table (Zipf, 1932, 1949). In simpler terms, the most frequently used word appears roughly twice as often as the second most frequent word, three times as often as the third most frequent word, and so on. If a dictionary includes a word that is significantly more frequent than other words in the

dictionary, it can greatly influence the analysis. This dominance can either be beneficial or detrimental, depending on the context. An instance where this proved advantageous was highlighted in Chapter 4, where I constructed a COVID-19 dictionary. At first, I compiled a relatively concise set of words to detect tweets related to the pandemic. Subsequently, I contrasted the outcomes with those obtained from a more exhaustive dictionary. Surprisingly, the performance of the concise dictionary was comparable to that of the comprehensive one, despite its omission of infrequent occurrences of certain words. Conversely, the inclusion of a frequently used word in a dictionary, which is erroneously categorized, can have significant disadvantageous implications, as demonstrated in the case of the words “tear” and “cry” in Study 3. These words comprised 2.7% of all tweets, highlighting the potential risks for false analysis and interpretations.

Modelling social media observational data poses challenges due to the data distribution patterns, which often include numerous extreme outliers and numerous zero values. On social media, most tweets receive no engagement, some garner moderate engagement, and relatively few become exceptionally viral. Dismissing any of these cases outright would be inappropriate. Viral tweets, for instance, are more than just numerical outliers as they naturally occur and hold practical significance, considering the large number of people who view and interact with content that went viral. Addressing the resulting left-skewed distribution containing zeros can be approached in various ways. In my thesis, I generally opted for log-modulus transformations as a consistent choice for data transformation. Nevertheless, alternative transformations, such as the reciprocal transformation denoted by $f(x) = 1-(x+1)^{-1}$ (see Appendix of Chapter 3), can also be employed. Additionally, alternative regression models that assume more suitable distributions for the untransformed data, such as binomial or Poisson regressions, have been employed by

other researchers (Brady et al., 2017; Rathje et al., 2021). Nevertheless, a comparison of these various methods in Studies 2-3 (see Appendix of Chapter 2) revealed similar outcomes. Model fit, evaluated by the AIC of the transformed data using both reciprocal and log-modulus transformations, consistently yielded higher values compared to the use of Poisson or Negative binomial regressions.

Lastly, it is important to note that the number of observations per individual can vary significantly. In Study 4, for instance, participants' tweet counts during the one-month observation period ranged from 1 to 600. To manage this wide range, I filtered the data to include only users with up to 200 tweets. Likewise, the number of tweets posted on any given day can exhibit variation. These nested structures can be effectively handled using multilevel regressions, which account for these variations by incorporating random intercepts or slopes.

Limitations and Future Directions

There are two types of limitations that I would like to emphasise in the context of this section. First, there are limitations of the above-mentioned findings regarding the assessment of affective content, the measurement of engagement, and the modelling of observational data derived from social media. Second, there are broader limitations of using the methods I chose to answer my research questions beyond the specific limitations I detailed in the previous two sections.

It is essential to note that the empirical work conducted in this thesis was not specifically designed to compare different methods for the assessment of affective content, the measurement of engagement, and the modelling of observational social media data. While I acquired insights about the efficiency of diverse techniques in certain contexts and for specific research inquiries, it's important to underline that my thesis studies were principally

designed to address two primary research questions concerning the prevalence and consequences of sharing, not to execute exhaustive comparative evaluations of research methodologies. Future research could strive to establish best practices by designing studies with a specific focus on contrasting different methodologies.

One limitation that underlies the empirical work of this thesis as a whole concern the interpretation of affective content. While the content produced by users does appear to relate to what we might expect them to be feeling—for instance, more negative affective content during crises and more positive affective content during political victories—it does not conclusively prove that expressions using emotional language accurately map onto individuals' actual emotions. Furthermore, although Study 4 reported in Chapter 4 identified similar patterns of affect expressions between Twitter users and participants in Gloria Wai Shan Ma's survey, it is important to note that the same individuals did not participate in both studies. In other words, I was still not able to directly link online expressions to self-reported emotions. Thus, the findings described in the studies of this thesis strictly pertain to the use of language and not to the experienced emotion. To address this limitation in future research, participants could be recruited to provide their social media handles (usernames to identify their accounts), along with regularly completing affect questionnaires such as the Positive and Negative Affect Schedule (PANAS, Watson, Clark, & Tellegen, 1988), in order to establish a connection between online expressions and subjective feeling.

An additional limitation of the observational studies in this thesis is their exclusive reliance on Twitter data. It is important to acknowledge that different social media platforms have distinct expression norms (Waterloo et al., 2018), user demographics (Bestvater, Shah, River, & Smith, 2022; Bossetta, 2018), and use purposes (Shearer & Grieco, 2019; Wood-Doughty, Smith, Broniatowski, & Dredze, 2017). The prevalent use of Twitter in research

can be attributed to the accessibility of its API, especially when compared to more restrictive platforms like Facebook or Weibo. However, this preference for Twitter-based research has led to the criticism of *Twitterology* where there is a risk of drawing conclusions that are specific to the platform itself. The risk of such a narrow focus on Twitter data is that such observational findings can often be transient, influenced by the ever-changing features and trends of the platform and the ephemeral nature of the internet (Bauckhage, 2011; Miltner, 2014). An example previously elaborated on in Chapter 2, is the fact that the usage and significance of hashtags on Twitter are subject to dynamic shifts (Booten, 2016). For instance, in Study 2, the hashtag #MAGA, initially associated primarily with conservatives, underwent a transformation over time, often being adopted ironically by liberals. To mitigate the impact of this evolving usage, data collection for the study was confined to a one-week period following election night. One way to alleviate such generalizability concerns is to ground observational findings in psychological theories that should apply across different contexts and platforms. For instance, in the case of Study 1, I could have conducted a descriptive analysis encompassing all potential characteristics that may or may not be linked to a higher likelihood of sharing negative affective content. However, the approach I defaulted to involved starting from a theoretical perspective, considering previous research that suggested tie-strength as one of the factors associated with increased sharing of negativity (Berger & Milkman, 2010; Fan et al., 2016). While the descriptive approach holds intrinsic informative value, particularly for those specifically interested in Twitter, the theoretical approach is arguably more likely to extend to multiple platforms beyond just Twitter, enhancing its applicability. An alternative option is to integrate both observational and experimental studies, as exemplified by the testing of the assumption regarding the influence of emotional language intensity on sharing in Studies 5 - 7. While it is noteworthy

that similar findings in regards to those presented in Study 2 - 3 have been reported by other researchers on other platforms (as discussed in Chapter 3, Fan et al., 2016; Rathje et al., 2021). The analysis of the content producer's role, as investigated in Study 1, or the effects of social sharing in Study 4, have yet to be replicated on other platforms. This presents an opportunity for future research to explore and replicate these findings in alternative platform settings.

The primary focus of the research presented in this thesis was on affective content within texts, inadvertently placing a potential constraint on the generalizability of the findings to affective text content, rather than images. As such, it remains an open question whether the patterns of sharing and the subsequent effects of exposure to image-based affective are similar to those of text-based content. Major social media platforms such as Instagram, YouTube, and TikTok heavily emphasize images and videos, while even on Twitter, which primarily centres around short text messages, images are frequently included. It is worth noting that positive images and videos, such as those featuring animals, have gained widespread popularity on the internet (Maddox, 2021), and have been shown to increase engagement in advertisements (Kujur & Singh, 2018) and political campaigns (Bossetta, 2018). Considering the findings of Study 5 – 8, which demonstrated that perceived emotional intensity is associated with increased sharing, it is reasonable to assume that not only emotionally intense text-based content but also emotionally intense images are likely to be shared more frequently. Nevertheless, it is crucial to recognize that the specific patterns of sharing behaviour may still differ between text and image content. The unique characteristics of each medium may contribute to potential variations in how users engage with and share content. Images can evoke immediate emotional responses and capture attention solely through visual processing (Nummenmaa, Hyönä, & Calvo, 2006). On the other hand, text-

based messages require users to engage in both visual and semantic processing, as they rely on the interpretation of written language (Derks et al., 2008; Harris & Paradice, 2007). These differences in processing requirements between images and text-based messages suggest that users may interact with and share content differently depending on the medium. This specific aspect was not explored within the scope of this thesis. Consequently, future research could delve into examining the differences and similarities between text and image content on social media platforms, in order to develop a more comprehensive understanding of content dissemination online.

Lastly, it is important to acknowledge that all types of engagement measures used in this thesis serve as proxies for actual content consumption. Users may consume vast amounts of content without leaving detectable engagement markers. Those who are most negatively impacted by social media are often passive consumers (Liu et al., 2019), and relying solely on measures of engagement such as likes or shares fails to capture their experiences. The experimental studies conducted in this research (Studies 5-8) aimed to provide insights into the content that people engage with, but it is crucial to note that these studies did not replicate an actual social media environment where individuals actively seek out content they prefer. In one recent study, Kelly and Sharot (2023) presented participants with Google search result pages and assessed which links they clicked on. Inspired by this work, a promising future approach could involve letting participants explore a social media platform while tracking their actual engagement. This could be accomplished by using mouse cursor movement as an indicator of attention and interaction (Johnson, Mulder, Sijbinga, & Hulsebos, 2012).

Broader Theoretical Considerations and Future Directions

In this thesis, my objective was to investigate the prevalence of affective content on social media by quantifying its creation and sharing and exploring the consequences of producing or viewing such content. In this final section, I aim to contextualize my findings within a broader framework and demonstrate that the insights gained from social media can be extrapolated to other contexts of affective transmission. Using this framework, this section endeavours to delineate crucial emerging research questions that ought to steer the future directions of the field.

Cultural Transmission Framework

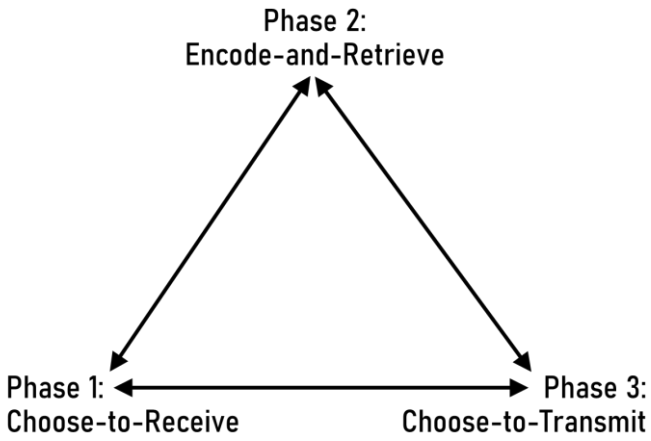
When we shift our focus away from social media and explore how emotional stories are disseminated among individuals offline, similarities between on- and offline transmissions become apparent in terms of the agents involved, behaviour exhibited, and the potential consequences that arise from the interaction. To illustrate, consider this situation: you are meeting a friend over a coffee, and they share the distressing news about the imminent closure of a local park due to a new real estate development. The park is dear to you and many others in the community, so you naturally feel inclined to discuss this issue with others in your social circle to rally support for the cause. This offline interaction mirrors a similar scenario that could occur on social media, where you follow a friend and come across their post about the park closure. To extend awareness about this local issue and galvanize community support, you share their post on your social media profile. In both instances, you are moved to share a negative news item with the aim to instigate collective action or awareness. This example highlights the overlap between online and offline emotional story sharing. Furthermore, the sense of emotional recovery may be quite similar

in both contexts as individuals find comfort and connection with fellow community members expressing shared concerns.

The cultural transmission model (CTM, see Figure 6.1), which aims to explain the transmission of cultural information and stories across generations, proposes a three-phase framework (for an overview, see K. Eriksson & Coultas, 2014)[15]. In Phase 1 (choose-to-receive), individuals actively choose whose stories they listen to and which ones they are interested in. Phase 2 (encode-and-retrieve) involves perceiving and later retelling a story, although, in the context of social media, this aspect is less relevant as stories can be easily shared without relying on memory. Finally, Phase 3 (choose-to-transmit) refers to the decision of whether and how to transmit a story to others (Heath, Bell, & Sternberg, 2001). Individuals make choices regarding the worthiness of the story for transmission and select appropriate methods for sharing it. These phases may not always occur in a strict linear sequence, and there exist bi-directional interactions between them. For example, consider the connection between Phase 1 and Phase 2. On one hand, our selection of individuals to listen to influences how we perceive the stories we encounter. On the other hand, the way we perceive stories can also shape our preferences and influence whom we choose to listen to. Each of these phases might be associated with different consequences for the sharing and perceiving parties.

[15] A similar three-phase model has recently also been proposed in the context of collective emotions (Goldenberg, 2022b)

Figure 6.1 *Cultural Transmission Model as Framework to Understand Affective Content on Social Media*



Note. *The primary emphasis of this thesis has been on investigating Phase 2 (Chapter 5) and Phase 3 (Chapter 2 – 4). However, it is important to acknowledge that Phase 1, which plays a vital role in determining the extent of users' exposure to affective content, requires further exploration. The acrophily literature suggests that our tendencies in network selection might also contribute to our exposure to such content.*

The research presented in this thesis has primarily focused on Phases 2 and 3 and the findings have three relevant implications. First, concerning Phase 2 (encode-and-retrieve), the gleaned insights regarding how individuals perceive sequential affective information in texts, as delineated in Chapter 5, may hold similar implications for verbal sequences. For example, if someone hears multiple people's opinions one after another during a discussion they might similarly overestimate the emotional intensity of the average opinion after the discussion. Second, findings regarding the creation and sharing of content add to the literature of Phase 3 (choose-to-transmit). Previous work using this framework has proposed similar tendencies to share more negativity as I did in Chapter 1, but my findings might suggest also adopting a

more nuanced perspective on negative affective content sharing. Drawing specifically from the findings of Studies 1 through 3, one could speculate that negative stories in a political context may generally be shared more widely, as indicated by Studies 2 and 3. In contrast, for non-political narratives, the strength of the tie between the story's transmitter and its recipient appears to sway the likelihood of either positive or negative stories being shared, as observed in Study 1.

Finally, this three-stage model exposes the lack of consideration of Phase 1 (choose-to-receive) throughout my thesis. Prior to users being exposed to other people's content, there are three influential factors shaping what individuals see. As already discussed in the *General Discussion* section, the first of these is the algorithm used to filter content that reaches users and the second is users' active information seeking. A third factor that operates even earlier in this process, is known as network selection. This factor revolves around the idea that users are exposed to high-affective content because they actively select individuals who have a tendency to produce such content. The emerging acrophily literature suggests that people tend to gravitate toward users who generate more intense content than they would create themselves (Goldenberg, Abruzzo, et al., 2022; Goldenberg et al., 2023; Saha et al., 2023). Consequently, users contribute not only to their own exposure to intense affective content online by creating and sharing such content but also by actively selecting other users who produce even more of this type of content.

Future Directions

There are three avenues of research that emerge as logical next steps in the investigation of online affective content after considering the work presented in this thesis.

First, following an examination of the characteristics of individuals whose negative affective content is most likely to be shared, future research should aim to uncover the traits of those who are most susceptible to sharing other users' negativity, thereby reiterating the distinction between producers and perceivers. Recent studies have indicated that only a small proportion of users are responsible for a large portion of negative affect sharing on social media platforms (Kumar et al., 2023), while the majority of other users do not actually approve of this type of content (Frimer et al., 2023). These types of users might have distinct features such as status-seeking and extreme political discontent, which leads them to toxic online behaviour (Petersen, Osmundsen, & Bor, 2021). Uncovering the motivations behind creating and sharing such content would be particularly intriguing for this line of inquiry.

Another important avenue of research is to examine mechanisms for sharing beyond the negativity bias. As discussed in the section on the *Prevalence of Positive and Negative Affective Content*, it is unlikely that the negativity bias universally drives sharing behaviour. A promising factor that could better explain the extent to which negative affective content is shared in specific contexts is the influence of emotional norms, as also proposed when examining the CTM (Tam, 2015). Generally speaking, behaviour is impacted by two types of norms, injunctive and descriptive norms (Berkowitz, 1997, 2005; Perkins & Berkowitz, 1986; Perkins, Craig, & Perkins, 2011). Injunctive norms refer to people's perception of what is considered acceptable by others. Descriptive norms refer to how others actually behave in general (Cialdini, Reno, & Kallgren, 1990). Both types of norms may play a significant role in shaping individuals' sharing behaviour, as they establish the accepted range of emotional expressions and reactions within a particular social or cultural setting. In contexts where negative affective expressions are perceived as more normative, such as in political contexts (Lau, 1985), people may feel licensed to express themselves negatively as well. Additionally,

indirect evidence for the influence of norms on sharing behaviour can be found in studies comparing sharing behaviour between cultures, which have found that a society's ideal affect influences society members' sharing behaviour (Hsu et al., 2021). Exploring the role of emotional norms and their impact not only provides better insights into when negative affective content is likely to be shared, but also offers an opportunity to develop interventions that can potentially influence the way people express themselves in specific contexts.

Arguably, the most important avenue to explore is the development of interventions to maximise the advantages and minimize the disadvantages of affective content expressions online. Here, researchers need to navigate both the ethical considerations and potential impacts of such interventions. Building upon the findings of this thesis, I propose two distinct types of interventions: one that harnesses the positive effects of social media and another that effectively reduces the negative effects. First, leveraging positive psychology to facilitate social interaction and foster feelings of connectedness through social media could be a promising approach (Regan et al., 2022). Empirical evidence indicates that extroverted users reap substantial benefits from social media (Bowden-Green et al., 2020); this inspired various positive psychology interventions that encourage individuals to adopt more extroverted behaviours (Rohrer, Richter, Brümmer, Wagner, & Schmukle, 2018), or establish social interaction goals (Reece et al., 2022; Schroeder, Lyons, & Epley, 2022), such as interacting with a specified number of people online to enhance individual well-being. Echoing this body of research, the results of Study 4 underscore the potential advantages of social media in harnessing a vast audience at all times. By emphasizing and promoting these aspects, we can enhance the positive effects of social media and create a sense of community and support among users. Second, while it has been observed that negative affective content might not be as prevalent as previously assumed, and the extent to which it can be deemed harmful is not

fully clear, there is less uncertainty regarding its prevalence and harm in the context of politics. In this realm, developing interventions that reduce the sharing of negative affective content may have the potential to alleviate political polarization and segregation. One potential strategy is the application of social norms interventions, which aim to modify people's behaviour by adjusting their perceptions of both descriptive and injunctive norms (Gimpel, Heger, Olenberger, & Utz, 2021). These interventions have proven efficacious in diverse areas, including bullying prevention (Berkowitz, 1997; Perkins et al., 2011), substance abuse (Neighbors, Larimer, & Lewis, 2004; Turner, Perkins, & Bauerle, 2008), and anti-littering campaigns (Cialdini et al., 1990; Schultz, Khazian, & Zaleski, 2008). In this particular case, an intervention might inform participants that negative affective content is neither well-received by the majority of users (an injunctive norm) nor particularly prevalent (a descriptive norm). By mitigating the negative impact of such content, we can contribute to a healthier online political discourse. Both types of interventions not only hold scientific merit but also have the potential to positively impact people's lives by fostering a healthier and more constructive social media environment.

Conclusion

Across eight reported studies, this thesis has aimed to improve our understanding of the prevalence of affective content on social media by quantifying its creation and sharing and exploring the consequences of producing or viewing such content. In doing so, this thesis has presented evidence indicating that positive affective content is generated more frequently than negative affective content. Furthermore, I have shown that negative affective content is shared more in political contexts and for users with large numbers of followers such as public figures. This suggests the existence of a general inclination towards positivity in content

creation, while sharing behaviour is influenced by both the specific context and the content sender. With regard to the consequences of affective content on social media, I found preliminary evidence that creating and sharing affective content could increase users' emotional well-being. On the other hand, viewing affective content in a sequential manner as is common on social media may lead to overestimation of the emotional language intensity of the discussed subjects as well as people's responses. Overall, I hoped this thesis could contribute to a more nuanced understanding of the prevalence of affective content online, as well as the mechanisms and consequences of its spread. This thesis thereby paves the way for utilizing communication technology to foster positive change for individuals' well-being and public discourse.

REFERENCES

Acerbi, A. (2019). *Cumulation. Cultural evolution in the digital age* (pp. 186 –212). Oxford University Press https//doi.org/10.1093/oso/9780198835943.003.0008

Adelman, J. S., & Estes, Z. (2013). Emotion and memory: A recognition advantage for positive and negative words independent of arousal. *Cognition*, *129*(3), 530-535.

Ahler, D. J., & Sood, G. (2018). The parties in our heads: Misperceptions about party composition and their consequences. *The Journal of Politics*, *80*(3), 964-981.

Albrecht, A. R., & Scholl, B. J. (2010). Perceptually averaging in a continuous visual world: Extracting statistical summary representations over time. *Psychological Science*, *21*(4), 560-567.

Alfasi, Y. (2019). The grass is always Barton, K. (2023). *MuMIn: Multi-Model Inference* (1.47.5).

greener on my Friends' profiles: The effect of Facebook social comparison on state self-esteem and depression. *Personality and Individual Differences*, *147*, 111-117.

Alvarez, R., Garcia, D., Moreno, Y., & Schweitzer, F. (2015). Sentiment cascades in the 15M movement. *EPJ Data Science*, *4*(1), 6.

Aral, S. (2020). *The hype machine: How social media disrupts our elections, our economy, and our health—and how we must adapt*. Currency.

Araujo, T., Wonneberger, A., Neijens, P., & de Vreese, C. (2017). How much time do you spend online? Understanding and improving the accuracy of self-reported measures of internet use. *Communication Methods and Measures*, *11*(3), 173-190.

Auxier, B. (2020). 64% of Americans say social media have a mostly negative effect on the way things are going in the U.S. today. *Pew Reseach Centre*. Retrieved from

https://www.pewresearch.org/fact-tank/2020/10/15/64-of-americans-say-social-media-have-a-mostly-negative-effect-on-the-way-things-are-going-in-the-u-s-today/

Bahrami, M., Findik, Y., Bozkaya, B., & Balcisoy, S. (2018). Twitter reveals: using twitter analytics to predict public protests. *arXiv preprint arXiv:1805.00358*. https://doi.org/10.48550/arXiv.1805.00358

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. F., . . . Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences, 115*(37), 9216-9221.

Barberá, P. (2015). Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. *Political Analysis, 23*(1), 76-91.

Barberá, P., Wang, N., Bonneau, R., Jost, J. T., Nagler, J., Tucker, J., & González-Bailón, S. (2015). The critical periphery in the growth of social protests. *PloS One, 10*(11), e0143611.

Barnett, A. G., Van Der Pols, J. C., & Dobson, A. J. (2005). Regression to the mean: what it is and how to deal with it. *International Journal of Epidemiology, 34*(1), 215-220.

Barton, K. (2023). MuMIn: Multi-Model Inference (1.47.5).

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. ArXiv Preprint ArXiv:1406.5823.

Bauckhage, C. (2011). Insights into internet memes. In Proceedings of the International AAAI Conference on Web and social Media (Vol. 5, No. 1, pp. 42-49).

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5*(4), 323-370.

Baylis, P., Obradovich, N., Kryvasheyeu, Y., Chen, H., Coviello, L., Moro, E., . . . Fowler, J. H. (2018). Weather impacts expressed sentiment. *PloS One, 13*(4), e0195750.

Bebbington, K., MacLeod, C., Ellison, T. M., Fay, N. (2017). The sky is falling: evidence of a negativity bias in the social transmission of information. *Evolution and Human Behavior*, *38*(1), 92-101.

Becerra, R., Preece, D., Campitelli, G., & Scott-Pillow, G. (2019). The assessment of emotional reactivity across negative and positive emotions: Development and validation of the Perth Emotional Reactivity Scale (PERS). *Assessment*, *26*(5), 867-879.

Bellovary, A. K., Young, N. A., & Goldenberg, A. (2021). Left-and right-leaning news organizations use negative emotional content and elicit user engagement similarly. *Affective Science, 2*(4), 391-396.

Berger, J., & Milkman, K. (2010). Social transmission, emotion, and the virality of online content. *Wharton Research Paper, 106*, 1-52.

Berger, J., & Packard, G. (2018). Are atypical things more popular? *Psychological Science*, *29*(7), 1178-1184.

Berkowitz, A. D. (1997). From reactive to proactive prevention: Promoting an ecology of health on campus. In P. C. Rivers & E. R. Shore (Eds.), *Substance abuse on campus: A handbook for college and university personnel* (pp. 119-139). Westport, CT: Greenwood Press.

Berkowitz, A. D. (2005). An overview of the social norms approach. *Changing the Culture of College Drinking: A Socially Situated Health Communication Campaign, 1*, 193-214.

Bernabé-Moreno, J., Tejeda-Lorente, A., Porcel, C., Fujita, H., & Herrera-Viedma, E. (2018). Quantifying the emotional impact of events on locations with social media. *Knowledge-Based Systems, 146*, 44-57.

Bestvater, S., Shah, S., River, G., & Smith, A. (2022). Politics on twitter: One-third of tweets from us adults are political. *Pew Research Centre.* https://www.pewresearch.org/politics/2022/06/16/ politics-on-twitter-one-third-of-tweets-from-u-s-adultsare-political/.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM, 55*(4), 77-84.

Blei, D. M., & Lafferty, J. (2009). Topic models. *Text Mining: Classification, Clustering, Applications, 10*(71), 34.

Bollen, J., & Mao, H. (2011). Twitter mood as a stock market predictor. *Computer, 44*(10), 91-94.

Bollen, J., Mao, H., & Pepe, A. (2011). Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. Proceedings of the International AAAI Conference on Web and Social Media, 5(1), 450–453.

Booten, K. (2016). Hashtag drift: Tracing the evolving uses of political hashtags over time. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, 2401–2405.

Bösch, K., Müller, O., & Schneider, J. (2018). Emotional Contagion Through Online Newspapers. In *ECIS proceedings* (p. 171).

Bossetta, M. (2018). The digital architectures of social media: Comparing political campaigning on Facebook, Twitter, Instagram, and Snapchat in the 2016 US election. *Journalism & Mass Communication Quarterly, 95*(2), 471-496.

Boutyline, A., & Willer, R. (2017). The social structure of political echo chambers: Variation in ideological homophily in online networks. *Political Psychology, 38*(3), 551-569.

Bowden-Green, T., Hinds, J., & Joinson, A. (2020). How is extraversion related to social media use? A literature review. *Personality and Individual Differences, 164*, 110040.

Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW): Instruction manual and affective ratings* (Vol. 30, No. 1, pp. 25-36). Technical report C-1, the center for research in psychophysiology, University of Florida.

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, *45*(4), 602–607.

Brady, W. J., Crockett, M., & Van Bavel, J. J. (2019). The MAD model of moral contagion: The role of motivation, attention and design in the spread of moralized content online. *Perspectives on Psychological Science, 15*(4), 978-1010.

Brady, W. J., & Crockett, M. J. (2019). How effective is online outrage? *Trends in Cognitive Sciences, 23*(2), 79-80.

Brady, W. J., Gantman, A. P., & Van Bavel, J. J. (2020). Attentional capture helps explain why moral and emotional content go viral. *Journal of Experimental Psychology: General, 149*(4), 746-756.

Brady, W. J., Jackson, J. C., Lindström, B., & Crockett, M. (2023). Algorithm-mediated social learning in online social networks. *OSF Preprints.* https://doi. org/10.31219/osf. io/yw5ah

Brady, W. J., McLoughlin, K., Doan, T. N., & Crockett, M. (2021). How social learning amplifies moral outrage expression in online social networks. *Science Advances, 7*(33).

Brady, W. J., McLoughlin, K., Torres, M., Luo, K., Gendron, M., & Crockett, M. (2022). Overperception of moral outrage in online social networks inflates beliefs about intergroup hostility. *Nature Human Behaviour,* 1-11.

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences, 114*(28), 7313-7318.

Brinberg, M., Ram, N., Yang, X., Cho, M.-J., Sundar, S. S., Robinson, T. N., & Reeves, B. (2021). The idiosyncrasies of everyday digital lives: Using the Human Screenome Project to study user behavior on smartphones. *Computers in Human Behavior, 114*, 106570.

Brown, J. R., & Enos, R. D. (2021). The measurement of partisan sorting for 180 million voters. *Nature Human Behaviour, 5*(8), 998-1008.

Brudner, E. G., Fareri, D. S., Shehata, S. G., & Delgado, M. R. (2022). Social feedback promotes positive social sharing, trust, and closeness. *Emotion*. Advance online publication. https://doi.org/10.1037/emo0001182

Buck, R., Losow, J. I., Murphy, M. M., & Costanzo, P. (1992). Social facilitation and inhibition of emotional expression and communication. *Journal of Personality and Social Psychology, 63*(6), 962.

Bunz, U. (2021). Investigating the relationship between social media use, big five personality, and well-being. *Journal of Communication Technology, 4*(3), 25-52.

Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin, 115*(3), 401.

Campbell, T., O'Brien, E., Van Boven, L., Schwarz, N., & Ubel, P. (2014). Too much experience: a desensitization bias in emotional perspective taking. *Journal of Personality and Social Psychology, 106*(2), 272.

Centola, D. (2010). The spread of behavior in an online social network experiment. *Science, 329*(5996), 1194-1197.

Pew Research Centre (2022). The political content in users' tweets and the accounts they follow. Retrieved from https://www.pewresearch.org/politics/2022/06/16/the-political-content-in-users-tweets-and-the-accounts-they-follow/

Cetina Presuel, R., & Martínez Sierra, J. M. (2019). Algorithms and the news: social media platforms as news publishers and distributors. *Revista De Comunicación, 18*(2), 261-285.

Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, P. K. (2010). Measuring user influence in twitter: The million follower fallacy. *ICWSM, 10*(30), 10-17.

Chang, D., & Ghim, G. (2011). The structure and dynamics of the Korean Twitter network. *Journal of Communication Research, 48*(1), 59-86.

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. *Advances in Neural Information Processing Systems, 22*.

Chau, M., & Xu, J. (2007). Mining communities and their relationships in blogs: A study of online hate groups. *International Journal of Human-Computer Studies, 65*(1), 57-70.

Choi, H., Oishi, S., Shin, J., & Suh, E. M. (2019). Do happy events love company? Cultural variations in sharing positive events with others. *Personality and Social Psychology Bulletin, 45*(4), 528-540.

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality Social Psychology, 58*(6), 1015.

Cikara, M. (2015). Intergroup schadenfreude: Motivating participation in collective violence. *Current Opinion in Behavioral Sciences, 3*, 12-17.

Citra, A. (2018). Maintaining the Good Image of a Public Figure through the use of Reputation Marketing. *Expose: Jurnal Ilmu Komunikasi*, *1*(1).

Cohen-Chen, S., Pliskin, R., & Goldenberg, A. (2020). Feel good or do good? A valence–function framework for understanding emotions. *Current Directions in Psychological Science, 29*(4), 388-393.

Collins, N. L., & Miller, L. C. (1994). Self-disclosure and liking: a meta-analytic review. *Psychological Bulletin, 116*(3), 457.

Crockett, M. J. (2017). Moral outrage in the digital age. *Nature Human Behaviour, 1*(11), 769-771.

Cummins, R. A., & Nistico, H. (2002). Maintaining life satisfaction: The role of positive cognitive bias. *Journal of Happiness Studies, 3*(1), 37.

de Mello, V. O., Cheung, F., & Inzlicht, M. (2022). Twitter use in the everyday life: exploring how Twitter use predicts well-being, polarization, and sense of belonging. *OSF Preprint*, https://doi.org/10.31234/osf.io/4x5em

De Rivera, J., & Páez, D. (2007). Emotional climate, human security, and cultures of peace. *Journal of Social Issues, 63*(2), 233-253.

DeLongis, A., & Holtzman, S. (2005). Coping in context: The role of stress, social support, and personality in coping. *Journal of Personality, 73*(6), 1633-1656.

Derks, D., Fischer, A. H., & Bos, A. E. (2008). The role of emotion in computer-mediated communication: A review. *Computers in Human Behavior, 24*(3), 766-785.

Derlega, V. J., Winstead, B. A., Wong, P. T., & Greenspan, M. (1987). Self-disclosure and relationship development: An attributional analysis. In M. E. Roloff & G. R. Miller (Eds.), *Interpersonal processes: New directions in communication research* (pp. 172–187). Thousand Oaks, CA: Sage.

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment, 49*(1), 71-75.

Diener, E., Heintzelman, S. J., Kushlev, K., Tay, L., Wirtz, D., Lutes, L. D., & Oishi, S. (2017). Findings all psychologists should know from the new science on subjective well-being. *Canadian Psychology, 58*(2), 87.

Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D. W., Oishi, S., & Biswas-Diener, R. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social Indicators Research, 97*(2), 143-156.

Dimitrov, D., Baran, E., Fafalios, P., Yu, R., Zhu, X., Zloch, M., & Dietze, S. (2020). Tweetscov19-a knowledge base of semantically annotated tweets about the covid-19 pandemic. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management* (pp. 2991-2998).

Dodds, P. S., Clark, E. M., Desu, S., Frank, M. R., Reagan, A. J., Williams, J. R., . . . Bagrow, J. P. (2015). Human language reveals a universal positivity bias. *Proceedings of the National Academy of Sciences, 112*(8), 2389-2394.

Doré, B., Ort, L., Braverman, O., & Ochsner, K. N. (2015). Sadness shifts to anxiety over time and distance from the national tragedy in Newtown, Connecticut. *Psychological Science, 26*(4), 363-373.

Duprez, C., Christophe, V., Rimé, B., Congard, A., & Antoine, P. (2015). Motives for the social sharing of an emotional experience. *Journal of Social and Personal Relationships, 32*(6), 757-787.

Eichstaedt, J. C., Smith, R. J., Merchant, R. M., Ungar, L. H., Crutchley, P., Preoţiuc-Pietro, D., . . . Schwartz, H. A. (2018). Facebook language predicts depression in medical records. *Proceedings of the National Academy of Sciences, 115*(44), 11203-11208.

Eimer, M., & Holmes, A. (2007). Event-related brain potential correlates of emotional face processing. *Neuropsychologia, 45*(1), 15-31.

Elfenbein, H. A., & Ambady, N. (2003). When familiarity breeds accuracy: cultural exposure and facial emotion recognition. *Journal of Personality Social Psychology, 85*(2), 276.

Elfenbein, H. A., Marsh, A. A., & Ambady, N. (2002). Emotional intelligence and the recognition of emotion from facial expressions. In L. F. Barrett & P. Salovey (Eds.), *The wisdom of feelings: Processes underlying emotional intelligence* (pp. 37–59). New York, NY: Guilford Press.

Eranti, V., & Lonkila, M. (2015). The social significance of the Facebook Like button. *First Monday, 20*(6).

Eriksson, K., & Coultas, J. C. (2014). Corpses, maggots, poodles and rats: emotional selection operating in three phases of cultural transmission of urban legends. *Journal of Cognition and Culture, 14*(1-2), 1-26.

Eriksson, M. (2016). Managing collective trauma on social media: The role of Twitter after the 2011 Norway attacks. *Media, Culture & Society, 38*(3), 365-380.

Fan, R., Xu, K., & Zhao, J. (2016). Higher contagion and weaker ties mean anger spreads faster than joy in social media. *arXiv preprint arXiv:1608.03656*.

Federico, C. M., Johnston, C. D., & Lavine, H. G. (2014). Context, engagement, and the (multiple) functions of negativity bias. *Behavioral and Brain Sciences, 37*(3), 311.

Feezell, J. T., Wagner, J. K., & Conroy, M. (2021). Exploring the effects of algorithm-driven news sources on political behavior and polarization. *Computers in Human Behavior, 116*, 106626.

Feinberg, M., Ford, B., Thai, S., Gatchpazian, A., & Lassetter, B. (2020). The political is personal: daily politics as a chronic stressor. *PsyArXiv*. https://doi.org/10 .31234/osf.io/hdz97

Ferrara, E., & Yang, Z. (2015). Measuring emotional contagion in social media. *PLoS One, 10*(11), e0142390. doi:10.1371/journal.pone.0142390

Ferrara, E., & Yang, Z. (2015). Quantifying the effect of sentiment on information diffusion in social media. *PeerJ Computer Science, 1*, e26.

Fine, J. A., & Hunt, M. F. (2021). Negativity and elite message diffusion on social media. *Political Behavior*, 1-19.

Finkel, E. J., Bail, C. A., Cikara, M., Ditto, P. H., Iyengar, S., Klar, S., . . . Rand, D. G. (2020). Political sectarianism in America. *Science, 370*(6516), 533-536.

Fischer, A., Halperin, E., Canetti, D., & Jasini, A. (2018). Why we hate. *Emotion Review, 10*(4), 309-320.

Fiske, S. T. (1980). Attention and weight in person perception: The impact of negative and extreme behavior. *Journal of Personality Social Psychology, 38*(6), 889.

Fitzpatrick, N. (2018). Media manipulation 2.0: the impact of social media on news, competition, and accuracy. *Athens Journal of Mass Media and Communications*, *4*(1), 45-61.

Frederick, S., & Loewenstein, G. (1999). Hedonic adaptation. In D. Kahneman, E. Diener, & N. Schwarz. *Well-Being: The foundations of hedonic psychology* (pp. 302-329). Russel Sage Foundation.

Frimer, J. A., Aujla, H., Feinberg, M., Skitka, L. J., Aquino, K., Eichstaedt, J. C., & Willer, R. (2023). Incivility is rising among American politicians on Twitter. *Social Psychological and Personality Science, 14*(2), 259-269.

Furlow, N. E. (2011). Find us on Facebook: How cause marketing has embraced social media. *Journal of Marketing Development and Competitiveness, 5*(6), 61-64.

Gable, S. L., Reis, H. T., Impett, E. A., & Asher, E. R. (2004). What do you do when things go right? The intrapersonal and interpersonal benefits of sharing positive events. *Journal of Personality Social Psychology, 87*(2), 228.

Gajarla, V., & Gupta, A. (2015). Emotion detection and sentiment analysis of images. *Georgia Institute of Technology*, 1-4.

Garcia, D., Garas, A., & Schweitzer, F. (2012). Positive words carry less information than negative words. *EPJ Data Science, 1*(1), 1-12.

Garcia, D., & Rimé, B. (2019). Collective emotions and social resilience in the digital traces after a terrorist attack. *Psychological Science, 30*(4), 617-628.

Gerbner, G., Gross, L., Morgan, M., Signorielli, N., & Shanahan, J. (2002). Growing up with television: Cultivation processes. In J. Bryant & D. Zillman (Eds.) *Media effects: Advances in theory and research* (2nd ed., pp. 53-78): Mahwah, NJ: Eribaum.

Gerlitz, C., & Helmond, A. (2013). The like economy: Social buttons and the data-intensive web. *New Media & Society, 15*(8), 1348-1365.

Gerson, J., Plagnol, A. C., & Corr, P. J. (2016). Subjective well-being and social media use: Do personality traits moderate the impact of social comparison on Facebook? *Computers in Human Behavior, 63*, 813-822.

Gîfu, D., & Cioca, M. (2014). Detecting emotions in comments on forums. *International Journal of Computers Communications & Control, 9*(6), 694-702.

Gilbert, C., & Hutto, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 8, No. 1, pp. 216-225).

Gimpel, H., Heger, S., Olenberger, C., & Utz, L. (2021). The effectiveness of social norms in fighting fake news on social media. *Journal of Management Information Systems, 38*(1), 196-221.

Goldenberg, A. (2022, October 7). Collective Emotion Regulation. *OSF Preprint.* https://doi.org/10.31219/osf.io/tzsy4.

Goldenberg, A. (2022, October 9). What makes groups emotional? *OSF Preprint.* https://doi.org/10.31219/osf.io/zfkgp

Goldenberg, A., Abruzzo, J. M., Huang, Z., Schöne, J., Bailey, D., Willer, R., . . . Gross, J. J. (2022). Homophily and acrophily as drivers of political segregation. *Nature Human Behaviour*, 1-12.

Goldenberg, A., Bailey, D., Muric, G., Ferrara, E., Schöne, J., Willer, R., . . . Gross, J. (2023, March 15). Homophily and Acrophily as Drivers of Political Segregation on Social Media. *OSF Preprint.* https://doi.org/10.31219/osf.io/cmx4p

Goldenberg, A., Garcia, D., Halperin, E., & Gross, J. J. (2020). Collective emotions. *Current Directions in Psychological Science, 29*(2), 154-160.

Goldenberg, A., Garcia, D., Halperin, E., Zaki, J., Kong, D., Golarai, G., & Gross, J. J. (2020). Beyond emotional similarity: The role of situation-specific motives. *Journal of Experimental Psychology: General, 149*(1), 138.

Goldenberg, A., & Gross, J. J. (2020). Digital emotion contagion. *Trends in Cognitive Sciences, 24*(4), 316-328.

Goldenberg, A., LaFollette, K., Huang, Z., Weisz, E., & Cikara, M. (2022, August 1). Judgment of Crowds as Emotional Increases with the Proportion of Black Faces. *OSF Preprint.* https://doi.org/10.31234/osf.io/ys65p

Goldenberg, A., Schöne, J., Huang, Z., Sweeny, T. D., Ong, D. C., Brady, T. F., . . . Gross, J. J. (2022). Amplification in the evaluation of multiple emotional expressions over time. *Nature Human Behaviour, 6*(10), 1408-1416.

Goldenberg, A., Sweeny, T. D., Shpigel, E., & Gross, J. J. (2020). Is this my group or not? The role of ensemble coding of emotional expressions in group categorization. *Journal of Experimental Psychology: General*, *149*(3), 445.

Goldenberg, A., Weisz, E., Sweeny, T., Cikara, M., & Gross, J. (2020). The crowd emotion amplification effect. *Psychological Science*, *32*(3), 437-450.

González-Bailón, S., & De Domenico, M. (2021). Bots are less central than verified accounts during contentious political events. *Proceedings of the National Academy of Sciences, 118*(11), e2013443118.

Gosling, S. D., Rentfrow, P. J., & Swann Jr, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*(6), 504-528.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, 424-438.

Green-Pedersen, C., Mortensen, P. B., & Thesen, G. (2017). The incumbency bonus revisited: Causes and consequences of media dominance. *British Journal of Political Science, 47*(1), 131-148.

Gross, J. H., & Johnson, K. T. (2016). Twitter taunts and tirades: Negative campaigning in the age of Trump. *PS: Political Science Politics, 49*(4), 748-754.

Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological Inquiry, 26*(1), 1-26.

Grün, B., Hornik, K., & Grün, M. B. (2020). Topicmodels: An R package for fitting topic models. *Journal of Statistical Software, 40*, 1-30.

Gruzd, A., Doiron, S., & Mai, P. (2011). Is happiness contagious online? A case of Twitter and the 2010 Winter Olympics. In *2011 44th Hawaii International Conference on System Sciences* (pp. 1-9). IEEE.

Guzman, E., Azócar, D., & Li, Y. (2014). Sentiment analysis of commit comments in GitHub: an empirical study. In *Proceedings of the 11th Working Conference on Mining Software Repositories* (pp. 352-355).

Haberman, J., Harp, T., & Whitney, D. (2009). Averaging facial expression over time. *Journal of Vision, 9*(11), 1-13.

Haberman, J., & Whitney, D. (2009). Seeing the mean: ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance, 35*(3), 718.

Hagen, S., Lochy, A., Jacques, C., Maillard, L., Colnat-Coulbois, S., Jonas, J., & Rossion, B. (2021). Dissociated face-and word-selective intracerebral responses in the human ventral occipito-temporal cortex. *Brain Structure and Function, 226*(9), 3031-3049.

Haggard, D. L., Robert, C., & Rose, A. J. (2011). Co-rumination in the workplace: Adjustment trade-offs for men and women who engage in excessive discussions of workplace problems. *Journal of Business and Psychology, 26*, 27-40.

Hajar, M. (2016). Using YouTube comments for text-based emotion recognition. *Procedia Computer Science, 83*, 292-299.

Halperin, E., Canetti, D., & Kimhi, S. (2012). In love with hatred: Rethinking the role hatred plays in shaping political behavior. *Journal of Applied Social Psychology, 42*(9), 2231-2256.

Hancock, J. T., Gee, K., Ciaccio, K., & Lin, J. M. H. (2008). *I'm sad you're sad: emotional contagion in CMC.* In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work* (pp. 295–298). ACM.

Hansen, L. K., Arvidsson, A., Nielsen, F. Å., Colleoni, E., & Etter, M. (2011). Good friends, bad news-affect and virality in twitter. In *Future Information Technology* (pp. 34-43): Springer.

Hardeniya, T., & Borikar, D. A. (2016). Dictionary based approach to sentiment analysis-a review. *International Journal of Advanced Engineering, Management and Science, 2*(5), 239438.

Harlow, R. E., & Cantor, N. (1995). To whom do people turn when things go poorly? Task orientation and functional social contacts. *Journal of Personality and Social Psychology, 69*(2), 329.

Harris, R. B., & Paradice, D. (2007). An investigation of the computer-mediated communication of emotions. *Journal of Applied Sciences Research, 3*(12), 2081-2090.

Hasan, M. R., Maliha, M., & Arifuzzaman, M. (2019). *Sentiment analysis with NLP on Twitter data.* In *2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2)* (pp. 1-4). IEEE.

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1994). *Emotional Contagion*. Cambridge University Press.

Haustein, S., Bowman, T. D., Holmberg, K., Tsou, A., Sugimoto, C. R., & Larivière, V. (2016). Tweets as impact indicators: Examining the implications of automated "bot" accounts on Twitter. *Journal of the Association for Information Science and Technology*, *67*(1), 232-238.

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Publications.

Headey, B., & Wearing, A. (1989). Personality, life events, and subjective well-being: Toward a dynamic equilibrium model. *Journal of Personality and Social Psychology*, *57*(4), 731.

Heath, C., Bell, C., & Sternberg, E. (2001). Emotional selection in memes: the case of urban legends. *Journal of Personality and Social Psychology*, *81*(6), 1028.

Heine, S. J., Lehman, D. R., Markus, H. R., & Kitayama, S. (1999). Is there a universal need for positive self-regard? *Psychological Review*, *106*(4), 766.

Hogan, B. (2010). The presentation of self in the age of social media: Distinguishing performances and exhibitions online. *Bulletin of Science, Technology & Society*, *30*(6), 377-386.

Hothorn, T., Zeileis, A., Farebrother, R. W., Cummins, C., Millo, G., Mitchell, D., & Zeileis, M. A. (2015). Package 'lmtest'. Testing linear regression models. https://cran.r-project. org/web/packages/lmtest/lmtest.pdf.

Hsu, T. W., Niiya, Y., Thelwall, M., Ko, M., Knutson, B., & Tsai, J. L. (2021). Social media users produce more affect that supports cultural values, but are more influenced by affect that violates cultural values. *Journal of Personality and Social Psychology, 121*(5), 969.

Hubert-Wallander, B., & Boynton, G. M. (2015). Not all summary statistics are made equal: Evidence from extracting summaries across time. *Journal of Vision, 15*(4), 5-5.

Huszár, F., Ktena, S. I., O'Brien, C., Belli, L., Schlaikjer, A., & Hardt, M. (2022). Algorithmic amplification of politics on Twitter. *Proceedings of the National Academy of Sciences, 119*(1), e2025334119.

Hyvärinen, H., & Beck, R. (2018). Emotions trump facts: The role of emotions in on social media: A literature review. In T. Bui (Ed.): *Proceedings of the Annual Hawaii International Conference on System Sciences, Proceedings of the 51st Hawaii International Conference on System Sciences*. Hawaii International Conference on System Sciences.

Iasiello, M., & Van Agteren, J. (2020). Mental health and/or mental illness: A scoping review of the evidence and implications of the dual-continua model of mental health. *Evidence Base, 1*, 1-45.

Imran, A. S., Daudpota, S. M., Kastrati, Z., & Batra, R. (2020). Cross-cultural polarity and emotion detection using sentiment analysis and deep learning on COVID-19 related tweets. *Ieee Access, 8*, 181074-181090.

Isen, A. M. (1984). Toward understanding the role of affect in cognition. In R. S. Wyer Jr. & T. K. Srull (Eds.), *Handbook of social cognition* (vol. 3, pp. 179-236). Hillsdale, NJ: Erlbaum.

Ito, T., Tsubouchi, K., Sakaji, H., Yamashita, T., & Izumi, K. (2020). Contextual sentiment neural network for document sentiment analysis. *Data Science and Engineering, 5*(2), 180-192.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science, 22*, 129-146.

Jacobi, C., Van Atteveldt, W., & Welbers, K. (2016). Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital Journalism, 4*(1), 89-106.

Jaidka, K., Giorgi, S., Schwartz, H. A., Kern, M. L., Ungar, L. H., & Eichstaedt, J. C. (2020). Estimating geographic subjective well-being from Twitter: A comparison of dictionary and data-driven language methods. *Proceedings of the National Academy of Sciences, 117*(19), 10165-10171.

Jalonen, H. (2014). Social media – an arena for venting negative emotions. *Online Journal of Communication and Media Technologies, 4*(October 2014-Special Issue), 53-70.

Johnson, A., Mulder, B., Sijbinga, A., & Hulsebos, L. (2012). Action as a window to perception: measuring attention with mouse movements. *Applied Cognitive Psychology, 26*(5), 802-809.

Jordan, J. J., Hoffman, M., Bloom, P., & Rand, D. G. (2016). Third-party punishment as a costly signal of trustworthiness. *Nature, 530*(7591), 473-476.

Jose, P. E., Lim, B. T., & Bryant, F. B. (2012). Does savoring increase happiness? A daily diary study. *The Journal of Positive Psychology, 7*(3), 176-187.

Kanyangara, P., Rimé, B., Philippot, P., & Yzerbyt, V. (2007). Collective rituals, emotional climate and intergroup perception: Participation in "Gacaca" tribunals and assimilation of the Rwandan genocide. *Journal of Social Issues, 63*(2), 387-403.

Kedrowicz, A. A., Royal, K., & Flammer, K. (2016). Social media and impression management: Veterinary medicine students' and faculty members' attitudes toward the acceptability of social media posts. *Journal of Advances in Medical Education Professionalism, 4*(4), 155.

Kee, Y. H., Li, C., Kong, L. C., Tang, C. J., & Chuang, K.-L. (2019). Scoping review of mindfulness research: A topic modelling approach. *Mindfulness*, 1-15.

Keipi, T., Räsänen, P., Oksanen, A., Hawdon, J., & Näsi, M. (2018). Exposure to online hate material and subjective well-being: A comparative study of American and Finnish youth. *Online Information Review*, *42*(1), 2-15.

Kelly, C., & Sharot, T. (2023). Knowledge-Seeking Reflects and Shapes Well-Being. *OSF Preprint.* https://doi.org/10.31234/osf.io/yd6j5

Keltner, D., Bowman, R., & Richards, H. (2017). *Exploring the emotional state of 'real happiness'. A study into the effects of watching natural history television content.* BBC.

Kensinger, E. A., & Corkin, S. (2003). Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words? *Memory Cognition, 31*(8), 1169-1180.

Keyes, C. L. (2007). Promoting and protecting mental health as flourishing: a complementary strategy for improving national mental health. *American Psychologist, 62*(2), 95.

Kim, H. S. (2009). Social sharing of emotion in words and otherwise. *Emotion Review*, *1*(1), 92-93.

Kissler, J., Herbert, C., Peyk, P., & Junghofer, M. (2007). Buzzwords: early cortical responses to emotional words during reading. *Psychological Science, 18*(6), 475-480.

Kloumann, I. M., Danforth, C. M., Harris, K. D., Bliss, C. A., & Dodds, P. S. (2012). Positivity of the English language. *PloS One, 7*(1), e29484.

Kraft, P. W., Krupnikov, Y., Milita, K., Ryan, J. B., & Soroka, S. (2020). Social media and the changing information environment: Sentiment differences in read versus recirculated news content. *Public Opinion Quarterly*, *84*(S1), 195-215.

Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, *111*(24), 8788-8790.

Kross, E., Verduyn, P., Sheppes, G., Costello, C. K., Jonides, J., & Ybarra, O. (2020). Social media and well-being: Pitfalls, progress, and next steps. *Trends in Cognitive Sciences*, *25*(1), 55-65.

Kujur, F., & Singh, S. (2018). Emotions as predictor for consumer engagement in YouTube advertisement. *Journal of Advances in Management Research*. *15*(2), 184-197.

Kumar, D., Hancock, J., Thomas, K., & Durumeric, Z. (2023, April). Understanding the behaviors of toxic accounts on reddit. In *Proceedings of the ACM Web Conference 2023* (pp. 2797-2807).

Kümpel, A. S., Karnowski, V., & Keyling, T. (2015). News sharing in social media: A review of current research on news sharing users, content, and networks. *Social Media & Society*, *14*(1), 1–14. doi:10.1177/2056305115610141

Kushawaha, D., De, D., Mohindru, V., & Gupta, A. K. (2020). Sentiment analysis and mood detection on an Android platform using machine learning integrated with Internet of Things. In *Proceedings of ICRIC 2019: Recent Innovations in Computing* (pp. 223-238). Springer International Publishing.

Langston, C. A. (1994). Capitalizing on and coping with daily-life events: Expressive responses to positive events. *Journal of Personality Social Psychology, 67*(6), 1112.

LaRose, R., Connolly, R., Lee, H., Li, K., & Hales, K. D. (2014). Connection overload? A cross cultural study of the consequences of social media connection. *Information Systems Management, 31*(1), 59-73.

Latzer, M., Hollnbuchner, K., Just, N., & Saurwein, F. (2016). The economics of algorithmic selection on the Internet. In *Handbook on the Economics of the Internet* (pp. 395-425). Edward Elgar Publishing.

Lau, R. R. (1985). Two explanations for negativity effects in political behavior. *American Journal of Political Science, 29*(1): 119–138.

Lazard, A. J., Collins, M. K. R., Hedrick, A., Varma, T., Love, B., Valle, C. G., . . . Benedict, C. (2021). Using social media for peer-to-peer cancer support: Interviews with young adults with cancer. *JMIR Cancer, 7*(3), e28234.

Lazarus, J. (2021). Negativity bias: An evolutionary hypothesis and an empirical programme. *Learning and Motivation, 75*, 101731.

Lee, H. R., Lee, H. E., Choi, J., Kim, J. H., & Han, H. L. (2014). Social media use, body image, and psychological well-being: A cross-cultural comparison of Korea and the United States. *Journal of Health Communication, 19*(12), 1343-1358.

Lees, J., & Cikara, M. (2020). Inaccurate group meta-perceptions drive negative out-group attributions in competitive contexts. *Nature Human Behaviour, 4*(3), 279-286.

Lengauer, G., Esser, F., & Berganza, R. (2012). Negativity in political news: A review of concepts, operationalizations and key findings. *Journalism, 13*(2), 179-202.

Lewallen, J., & Behm-Morawitz, E. (2016). Pinterest or thinterest?: Social comparison and body image on social media. *Social Media+ Society, 2*(1), 2056305116640559.

Li, H. (2018). Deep learning for natural language processing: advantages and challenges. *National Science Review, 5*(1), 24-26.

Li, J., Conathan, D., & Hughes, C. (2017, July). Rethinking emotional desensitization to violence: Methodological and theoretical insights from social media data. In *Proceedings of the 8th International Conference on Social Media & Society* (pp. 1-5).

Li, J., Wang, X., & Hovy, E. (2014, November). What a nasty day: Exploring mood-weather relationship from twitter. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management* (pp. 1309-1318).

Lin, H., Tov, W., & Qiu, L. (2014). Emotional disclosure on social networking sites: The role of network structure and psychological needs. *Computers in Human Behavior, 41*, 342-350.

Lindström, B., Bellander, M., Schultner, D. T., Chang, A., Tobler, P. N., & Amodio, D. M. (2021). A computational reward learning account of social media engagement. *Nature Communications, 12*(1), 1311.

Liu, D., Baumeister, R. F., Yang, C. C., & Hu, B. (2019). Digital communication media use and psychological well-being: A meta-analysis. *Journal of Computer-Mediated Communication, 24*(5), 259-273.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*(6657), 279-281.

Luminet IV, O., Bouts, P., Delie, F., Manstead, A. S., & Rimé, B. (2000). Social sharing of emotion following exposure to a negatively valenced situation. *Cognition & Emotion, 14*(5), 661-688.

Luo, M., & Hancock, J. T. (2020). Self-disclosure and social media: motivations, mechanisms and psychological well-being. *Current Opinion in Psychology, 31*, 110-115.

Ma, G. W. S., Schöne, J., & Parkinson, B. (2022). Social sharing of emotion during the collective crisis of COVID-19. *OSF Preprint.* https://doi.org/10.31219/osf.io/9p3wh

Maddox, J. (2021). The secret life of pet Instagram accounts: Joy, resistance, and commodification in the Internet's cute economy. *New Media & Society, 23*(11), 3332-3348.

Manikonda, L., Beigi, G., Kambhampati, S., & Liu, H. (2018). # metoo through the lens of social media. In *Social, Cultural, and Behavioral Modeling: 11th International Conference, SBP-BRiMS 2018, Washington, DC, USA, July 10-13, 2018, Proceedings 11* (pp. 104-110). Springer International Publishing.

Mannell, K., & Meese, J. (2022). From doom-scrolling to news avoidance: limiting news as a wellbeing strategy during COVID lockdown. *Journalism Studies, 23*(3), 302-319.

Manstead, A. S., & Fischer, A. H. (2001). Social appraisal: the social world as object of and influence on appraisal processes. In *Appraisal Processes in Emotion: Theory, Research, Application*, ed. KR Scherer, A Schorr, T Johnstone, pp. 221–32. New York: Oxford Univ. Press

Marroquín, B. (2011). Interpersonal emotion regulation as a mechanism of social support in depression. *Clinical Psychology Review, 31*(8), 1276-1290.

Martins, A., Ramalho, N., & Morin, E. (2010). A comprehensive meta-analysis of the relationship between emotional intelligence and health. *Personality and Individual Differences, 49*(6), 554-564.

Mattick, R. P., & Clarke, J. C. (1998). Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behaviour Research Therapy, 36*(4), 455-470.

Mauss, I. B., Shallcross, A. J., Troy, A. S., John, O. P., Ferrer, E., Wilhelm, F. H., & Gross, J. J. (2011). Don't hide your happiness! Positive emotion dissociation, social connectedness, and psychological functioning. *Journal of Personality and Social Psychology, 100*(4), 738.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology, 27*(1), 415-444.

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal, 5*(4), 1093-1113.

Meffert, M. F., Chung, S., Joiner, A. J., Waks, L., & Garst, J. (2006). The effects of negativity and motivated information processing during a political campaign. *Journal of Communication, 56*(1), 27-51.

Mercadante, E. J., Tracy, J. L., & Götz, F. M. (2023). Greed communication predicts the approval and reach of US senators' tweets. *Proceedings of the National Academy of Sciences, 120*(11), e2218680120.

Milkman, K. L., & Berger, J. (2014). The science of sharing and the sharing of science. *Proceedings of the National Academy of Sciences, 111*(4), 13642-13649.

Milli, S., Carroll, M., Pandey, S., Wang, Y., & Dragan, A. D. (2023). Twitter's algorithm: Amplifying anger, animosity, and affective polarization. *arXiv preprint arXiv:2305.16941*.

Miltner, K. M. (2014). “There’s no place for lulz on LOLCats”: The role of genre, gender, and group identity in the interpretation and enjoyment of an Internet meme. *First Monday*. *19*(8)

Mitchell, A., & Page, D. (2014). State of the news media. *Pew Research Center*. http://www.journalism.org/2014/03/26/state-of-the-news-media-2014-overview/.

Mitchell, G. W., & Glicksman, A. S. (1977). Cancer patients: Knowledge and attitudes. *Cancer*, *40*(1), 61-66.

Mitchell, L., Frank, M. R., Harris, K. D., Dodds, P. S., & Danforth, C. M. (2013). The geography of happiness: Connecting twitter sentiment and expression, demographics, and objective characteristics of place. *PloS One*, *8*(5), e64417.

Mohammad, S. M., & Turney, P. D. (2013). Nrc emotion lexicon. *National Research Council, Canada*, *2*, 234.

Müller, K., & Schwarz, C. (2021). Fanning the flames of hate: Social media and hate crime. *Journal of European Economic Association*, *19*(4), 2131-2167.

Mundt, M., Ross, K., & Burnett, C. M. (2018). Scaling social movements through social media: The case of Black Lives Matter. *Social Media+ Society*, *4*(4), 2056305118807911.

Nahar, L., Sultana, Z., Iqbal, N., & Chowdhury, A. (2019, May). Sentiment analysis and emotion extraction: A review of research paradigm. In *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)* (pp. 1-8). *IEEE*.

Nakagawa, S., Johnson, P. C., & Schielzeth, H. (2017). The coefficient of determination R 2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface, 14*(134), 20170213.

Näsi, M., Räsänen, P., Hawdon, J., Holkeri, E., Oksanen, A., & People. (2015). Exposure to online hate material and social trust among Finnish youth. *Information Technology, 28*(3), 607-622.

Neighbors, C., Larimer, M. E., & Lewis, M. A. (2004). Targeting misperceptions of descriptive drinking norms: efficacy of a computer-delivered personalized normative feedback intervention. *Journal of Consulting Clinical Psychology, 72*(3), 434.

Nelson, S. K., Della Porta, M. D., Jacobs Bao, K., Lee, H. C., Choi, I., & Lyubomirsky, S. (2015). 'It's up to you': Experimentally manipulated autonomy support for prosocial behavior improves well-being in two cultures over six weeks. *The Journal of Positive Psychology, 10*(5), 463-476.

Nemes, L., & Kiss, A. (2021). Social media sentiment analysis based on COVID-19. *Journal of Information and Telecommunication, 5*(1), 1-15.

Newson, M., Buhrmester, M., & Whitehouse, H. (2023). United in defeat: shared suffering and group bonding among football fans. *Managing Sport and Leisure, 28*(2), 164-181.

Nguyen, A., Glück, A., & Jackson, D. (2022). The up-down-up pandemic news experience: A mixed-method approach to its negative and positive effects on psychological wellbeing. *Journalism*, *0*(0). https://doi.org/10.1177/14648849221135137

Nummenmaa, L., Hyönä, J., & Calvo, M. G. (2006). Eye movement assessment of selective attentional capture by emotional pictures. *Emotion, 6*(2), 257.

O'Brien, E., & Kassirer, S. (2019). People are slow to adapt to the warm glow of giving. *Psychological Science, 30*(2), 193-204.

Okango, E., & Mwambi, H. (2022). Dictionary based global twitter sentiment analysis of coronavirus (COVID-19) effects and response. *Annals of Data Science, 9*(1), 175-186.

Orben, A., Przybylski, A. K., Blakemore, S.-J., & Kievit, R. A. (2022). Windows of developmental sensitivity to social media. *Nature Communications, 13*(1), 1649.

Otto, A. R., & Eichstaedt, J. C. (2018). Real-world unexpected outcomes predict city-level mood states and risk-taking behavior. *PloS One, 13*(11), e0206923.

Ottoni, R., Las Casas, D., Pesce, J. P., Meira Jr., W., Wilson, C., Mislove, A., & Almeida, V. (2014). Of Pins and Tweets: Investigating How Users Behave Across Image- and Text-Based Social Networks. *Proceedings of the International AAAI Conference on Web and Social Media, 8*(1), 386-395. https://doi.org/10.1609/icwsm.v8i1.14514

Páez, D., Rimé, B., Basabe, N., Wlodarczyk, A., & Zumeta, L. (2015). Psychosocial effects of perceived emotional synchrony in collective gatherings. *Journal of Personality and Social Psychology, 108*(5), 711.

Palmer, C. A., Ramsey, M. A., Morey, J. N., & Gentzler, A. L. (2016). How do people share their positive events? *Journal of Individual Differences*.

Park, S. Y., & Baek, Y. M. (2018). Two faces of social comparison on Facebook: The interplay between social comparison orientation, emotions, and psychological well-being. *Computers in Human Behavior, 79*, 83-93.

Parkinson, B., & Simons, G. (2012). Worry spreads: Interpersonal transfer of problem-related anxiety. *Cognition & Emotion, 26*(3), 462-479.

Parsons, C. A., Alden, L. E., & Biesanz, J. C. (2021). Influencing emotion: Social anxiety and comparisons on Instagram. *Emotion, 21*(7), 1427.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates, 71*(2001), 2001.

Perkins, H. W., & Berkowitz, A. D. (1986). Perceiving the community norms of alcohol use among students: Some research implications for campus alcohol education programming. *International Journal of the Addictions, 21*(9-10), 961-976.

Perkins, H. W., Craig, D. W., & Perkins, J. M. (2011). Using social norms to reduce bullying: A research intervention among adolescents in five middle schools. *Group Processes Intergroup Relations, 14*(5), 703-722.

Perrin, A. (2015). Social media usage. *Pew Research Center, 125*, 52-68. https://www.pewresearch.org/internet/2015/10/08/social-networking-usage-2005-2015/

Petersen, M. B., Osmundsen, M., & Bor, A. (2021). Beyond populism: The psychology of status-seeking and extreme political discontent. In *The Psychology of Populism* (pp. 62-80). Routledge.

Peterson, E., Goel, S., & Iyengar, S. (2018). Echo chambers and partisan polarization: Evidence from the 2016 presidential campaign. *Unpublished Manuscript. https://5harad.com/papers/selecfive-exposure/pdf*.

Peterson, S. (1981). International news selection by the elite press: A case study. *Public Opinion Quarterly, 45*(2), 143-163.

Petrov, A., & Proncheva, O. (2018). Modeling propaganda battle: decision-making, homophily, and echo chambers. In *Artificial Intelligence and Natural Language: 7th International Conference, AINL 2018, St. Petersburg, Russia, October 17–19, 2018, Proceedings* (pp. 197-209). Cham: Springer International Publishing.

Potts, R., & Sanchez, D. (1994). Television viewing and depression: No news is good news. *Journal of Broadcasting & Electronic Media, 38*(1), 79-90.

Rathje, S., Robertson, C., Brady, W. J., & Van Bavel, J. J. (2022). People think that social media platforms do (but should not) amplify divisive content. *PsyArXix Preprint*. https://doi.org/10.31234/osf.io/gmun4

Rathje, S., Van Bavel, J. J., & van der Linden, S. (2021). Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences, 118*(26).

Reece, A., Cooney, G., Bull, P., Chung, C., Dawson, B., Fitzpatrick, C., . . . Marin, S. (2022). Advancing an interdisciplinary science of conversation: insights from a large multimodal corpus of human speech. *arXiv Preprint. arXiv:2203.00674*.

Regan, A., Radošić, N., & Lyubomirsky, S. (2022). Experimental effects of social behavior on well-being. *Trends in Cognitive Sciences*. 26(11), 987-998

Reser, J. P., & Swim, J. K. (2011). Adapting to and coping with the threat and impacts of climate change. *American Psychologist, 66*(4), 277.

Ribeiro, F. N., Araújo, M., Gonçalves, P., Gonçalves, M. A., & Benevenuto, F. (2016). Sentibench-a benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Science, 5*(1), 1-29.

Rimé, B. (2007). The social sharing of emotion as an interface between individual and collective processes in the construction of emotional climates. *Journal of Social Issues, 63*(2), 307-322.

Rimé, B. (2009). Emotion elicits the social sharing of emotion: Theory and empirical review. *Emotion Review, 1*(1), 60-85.

Rimé, B., Bouchat, P., Paquot, L., & Giglio, L. (2020). Intrapersonal, interpersonal, and social outcomes of the social sharing of emotion. *Current Opinion in Psychology, 31*, 127-134.

Rimé, B., Finkenauer, C., Luminet, O., Zech, E., & Philippot, P. (1998). Social sharing of emotion: New evidence and new questions. *European Review of Social Psychology, 9*(1), 145-189.

Rimé, B., Páez, D., Basabe, N., & Martínez, F. (2010). Social sharing of emotion, post-traumatic growth, and emotional climate: Follow-up of Spanish citizen's response to the collective trauma of March 11th terrorist attacks in Madrid. *European Journal of Social Psychology, 40*(6), 1029-1045.

Roberts, J. A., & David, M. E. (2020). The social media party: Fear of missing out (FoMO), social media intensity, connection, and well-being. *International Journal of Human–Computer Interaction, 36*(4), 386-392.

Robertson, C. E., Pröllochs, N., Schwarzenegger, K., Parnamets, P., Van Bavel, J. J., & Feuerriegel, S. (2022). Negativity drives online news consumption. *Nature Human Behaviour*. *7*(5), 812-822.

Robinson, R. J., Keltner, D., Ward, A., & Ross, L. (1995). Actual versus assumed differences in construal:" Naive realism" in intergroup perception and conflict. *Journal of Personality and Social Psychology, 68*(3), 404.

Rocklage, M. D., Rucker, D. D., & Nordgren, L. F. (2018). The Evaluative Lexicon 2.0: The measurement of emotionality, extremity, and valence in language. *Behavior Research Methods, 50*(4), 1327-1344.

Rohrer, J. M., Richter, D., Brümmer, M., Wagner, G. G., & Schmukle, S. C. (2018). Successfully striving for happiness: Socially engaged pursuits predict increases in life satisfaction. *Psychological Science, 29*(8), 1291-1298.

Rose, A. J. (2002). Co–rumination in the friendships of girls and boys. *Child Development, 73*(6), 1830-1843.

Rose, A. J. (2021). The costs and benefits of co-rumination. *Child Development Perspectives, 15*(3), 176-181.

Rose, A. J., Carlson, W., & Waller, E. M. (2007). Prospective associations of co-rumination with friendship and emotional adjustment: considering the socioemotional trade-offs of co-rumination. *Developmental Psychology, 43*(4), 1019.

Ross, K., & Bürger, T. (2014). Face to face (book) Social media, political campaigning and the unbearable lightness of being there. *Political Science, 66*(1), 46-62.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality Social Psychology Review, 5*(4), 296-320.

Petersen, M. B., Osmundsen, M., & Bor, A. (2021). Beyond populism: The psychology of status-seeking and extreme political discontent. In *The Psychology of Populism* (pp. 62-80). Routledge.

Saha, P., Garimella, K., Kalyan, N. K., Pandey, S. K., Meher, P. M., Mathew, B., & Mukherjee, A. (2023). On the rise of fear speech in online social media. *Proceedings of the National Academy of Sciences, 120*(11), e2212270120.

Samal, B., Behera, A. K., & Panda, M. (2017, May). Performance analysis of supervised machine learning techniques for sentiment analysis. In *2017 Third International Conference on Sensing, Signal Processing and Security (ICSSS)* (pp. 128-133). IEEE.

Schoenberg, B., Carr, A. C., Peretz, D., Kutscher, A. H., & Cherico, D. J. (1975). Advice of the bereaved for the bereaved. In B. Schoenberg, I. Gerber, A. Wiener, AH Kutscher, D. Perets, D., & AC Carr *(Eds.), Bereavement: Its Psychosocial Aspects*, 362-367. New York: Columbia University Press.

Schöne, J., Garcia, D., Parkinson, B., & Goldenberg, A. (in press). Negative Expressions are Shared more on Twitter for Public Figures than for Ordinary Users. *PNAS Nexus.* https://doi.org/10.31234/osf.io/wng5v

Schöne, J., Parkinson, B., & Goldenberg, A. (2021). Negativity Spreads More than Positivity on Twitter after both Positive and Negative Political Situations. *Affective Science, 2*(4), 379-390. doi:10.31234/osf.io/x9e7u

Schroeder, J., Lyons, D., & Epley, N. (2022). Hello, stranger? Pleasant conversations are preceded by concerns about starting one. *Journal of Experimental Psychology: General, 151*(5), 1141.

Schultz, W. P., Khazian, A. M., & Zaleski, A. C. (2008). Using normative social influence to promote conservation among hotel guests. *Social Influence, 3*(1), 4-23.

Schuster, M. A., Stein, B. D., Jaycox, L. H., Collins, R. L., Marshall, G. N., Elliott, M. N., . . . Berry, S. H. (2001). A national survey of stress reactions after the September 11, 2001, terrorist attacks. *New England Journal of Medicine, 345*(20), 1507-1512.

Seidlitz, L., & Diener, E. (1993). Memory for positive versus negative life events: theories for the differences between happy and unhappy persons. *Journal of Personality Social Psychology, 64*(4), 654.

Shearer, E., & Grieco, E. (2019). Americans are wary of the role social media sites play in delivering the news. *Pew Research Center,*

https://www.pewresearch.org/journalism/2019/10/02/americans-are-wary-of-the-role-social-media-sites-play-in-delivering-the-news/

Silge, J., & Robinson, D. (2017). *Text mining with R: A tidy approach*. Boston, MA: O'Reilly Media.

Sillars, A., & Zorn, T. E. (2021). Hypernegative interpretation of negatively perceived email at work. *Management Communication Quarterly, 35*(2), 171-200.

Singh, V., Piryani, R., Uddin, A., Waila, P., & Marisha (2013). Sentiment analysis of textual reviews; Evaluating machine learning, unsupervised and SentiWordNet approaches. In *2013 5th International Conference on Knowledge and Smart Technology (KST)*, (pp. 122-127).

Skowronski, J. J., & Carlston, D. E. (1987). Social judgment and social memory: The role of cue diagnosticity in negativity, positivity, and extremity biases. *Journal of Personality and Social Psychology, 52*(4), 689.

Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin, 105*(1), 131.

Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A. Y., & Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing* (pp. 1631-1642).

Soroka, S. (2016). Gatekeeping and the negativity bias. In *Oxford Research Encyclopedia of Politics*.

Soroka, S., Fournier, P., & Nir, L. (2019). Cross-national evidence of a negativity bias in psychophysiological reactions to news. *Proceedings of the National Academy of Sciences, 116*(38), 18888-18892.

Soroka, S., & McAdams, S. (2015). News, politics, and negativity. *Political Communication, 32*(1), 1-22.

Sousa, M. G., Sakiyama, K., de Souza Rodrigues, L., Moraes, P. H., Fernandes, E. R., & Matsubara, E. T. (2019, November). BERT for stock market sentiment analysis. In *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)* (pp. 1597-1601). IEEE.

Spendelow, J. S., Simonds, L. M., & Avery, R. E. (2017). The relationship between co-rumination and internalizing problems: A systematic review and meta-analysis. *Clinical Psychology & Psychotherapy, 24*(2), 512-527.

Clement, J. (2022). Daily social media usage worldwide. *Statista* https://www.statista.com/statistics/433871/daily-social-media-usage-worldwide/.

Clement, J. (2022). Number of social network users worldwide from 2017 to 2025. *Statista* https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/.

Steers, M.-L. N., Wickham, R. E., & Acitelli, L. K. (2014). Seeing everyone else's highlight reels: How Facebook usage is linked to depressive symptoms. *Journal of Social Clinical Psychology, 33*(8), 701.

Steinert, S., & Dennis, M. J. (2022). Emotions and Digital Well-Being: on Social Media's Emotional Affordances. *Philosophy Technology, 35*(2), 1-21.

Stevens, H. R., Oh, Y. J., & Taylor, L. D. (2021). Desensitization to fear-inducing COVID-19 health news on Twitter: observational study. *JMIR infodemiology, 1*(1), e26876.

Stewart-Brown, S., & Janmohamed, K. (2008). Warwick-Edinburgh mental well-being scale. *User Guide. Version, 1*(10.1037).

Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior. *Journal of Management Information Systems, 29*(4), 217-248.

Stone, P. J., Dunphy, D. C., & Smith, M. S. (1966). The general inquirer: A computer approach to content analysis. M.I.T. Press.

Straughan, D. M. (1989). An experiment on the relation between news values and reader interest. *International Communication Gazette, 43*(2), 93-107.

Stuart, E. A., King, G., Imai, K., & Ho, D. (2011). MatchIt: nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*. *42*(8)

Subrahmanian, V. S., Azaria, A., Durst, S., Kagan, V., Galstyan, A., Lerman, K., . . . Menczer, F. (2016). The DARPA Twitter bot challenge. *Computer, 49*(6), 38-46.

Tam, K.-P. (2015). Understanding intergenerational cultural transmission through the role of perceived norms. *Journal of Cross-Cultural Psychology, 46*(10), 1260-1266.

Tamir, D. I., & Mitchell, J. P. (2012). Disclosing information about the self is intrinsically rewarding. *Proceedings of the National Academy of Sciences, 109*(21), 8038-8043.

Tamir, M. (2009). What do people want to feel and why? Pleasure and utility in emotion regulation. *Current Directions in Psychological Science, 18*(2), 101-105.

Tandoc Jr, E. C., Ferrucci, P., & Duffy, M. (2015). Facebook use, envy, and depression among college students: Is facebooking depressing? *Computers in Human Behavior, 43*, 139-146.

Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology, 61*(12), 2544-2558.

Tong, K., Dubé, C., & Sekuler, R. (2019). What makes a prototype a prototype? Averaging visual features in a sequence. *Attention, Perception, Psychophysics, 81*(6), 1962-1978.

Tsai, J. (2019). 15.2 Culture and Emotion. *Introduction to Psychology*.

Tsugawa, S., & Ohsaki, H. (2015, November). Negative messages spread rapidly and widely on social media. In *Proceedings of the 2015 ACM on conference on online social networks* (pp. 151-160).

Tucker, J. A., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., . . . Nyhan, B. (2018). Social media, political polarization, and political disinformation: A review of the scientific literature. *Political polarization, and political disinformation: a review of the scientific literature (March 19, 2018). SSRN Electron*.

Tufekci, Z. (2014, May). Big questions for social media big data: Representativeness, validity and other methodological pitfalls. In *Proceedings of the international AAAI conference on web and social media* (Vol. 8, No. 1, pp. 505-514).

Tufekci, Z. (2017). *Twitter and tear gas: The power and fragility of networked protest*. Yale University Press.

Tugade, M. M., & Fredrickson, B. L. (2007). Regulation of Positive Emotions. *Journal of Happiness Studies, 8*, 311-333.

Turnbull, S., & Jenkins, S. (2016). Why Facebook Reactions are good news for evaluating social media campaigns. *Journal of Direct, Data and Digital Marketing Practice, 17*, 156-158.

Turner, J., Perkins, H. W., & Bauerle, J. (2008). Declining negative consequences related to alcohol misuse among students exposed to a social norms marketing intervention on a college campus. *Journal of American College Health, 57*(1), 85-94.

Uhlir, J. L. (2016). Social comparison and self-presentation on social media as predictors of depressive symptoms. (Scripps Senior Theses, 756) [Thesis, Scripps College]. https://scholarship.claremont.edu/scripps_theses/756 .

van der Meer, T. G., & Hameleers, M. (2022). I Knew It, the World is Falling Apart! Combatting a Confirmatory Negativity Bias in Audiences' News Selection Through News Media Literacy Interventions. *Digital Journalism, 10*(3), 473-492.

Verduyn, P., Lee, D. S., Park, J., Shablack, H., Orvell, A., Bayer, J., . . . Kross, E. (2015). Passive Facebook usage undermines affective well-being: Experimental and longitudinal evidence. *Journal of Experimental Psychology: General, 144*(2), 480.

Vermeulen, A., Vandebosch, H., & Heirman, W. (2018). # Smiling,# venting, or both? Adolescents' social sharing of emotions on social media. *Computers in Human Behavior, 84*, 211-219.

Vogel, E. A., & Rose, J. P. (2016). Self-reflection and interpersonal connection: Making the most of self-presentation on social media. *Translational Issues in Psychological Science, 2*(3), 294.

Vogel, E. A., Rose, J. P., Okdie, B. M., Eckles, K., & Franz, B. (2015). Who compares and despairs? The effect of social comparison orientation on social media use and its outcomes. *Personality and Individual Differences, 86*, 249-256.

Wade, J. (2017). Social Media Week London 2017 Recap. Retrieved January 12, 2018.

Wang, H., Lei, K., & Xu, K. (2015, June). Profiling the followers of the most influential and verified users on Sina Weibo. In *2015 IEEE International Conference on Communications (ICC)* (pp. 1158-1163). IEEE.

Wang, Y., & Li, B. (2015, November). Sentiment analysis for social media images. In *2015 IEEE international conference on data mining workshop (ICDMW)* (pp. 1584-1591). IEEE.

Wang, Z., Hale, S., Adelani, D. I., Grabowicz, P., Hartman, T., & Jurgens, D. (2019, May). Demographic inference and representative population estimates from multilingual social media data. In *The world wide web conference* (pp. 2056-2067).

Waterloo, S. F., Baumgartner, S. E., Peter, J., & Valkenburg, P. M. (2018). Norms of online expressions of emotion: Comparing Facebook, Twitter, Instagram, and WhatsApp. *New Media Society, 20*(5), 1813-1831.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality Social Psychology, 54*(6), 1063.

Weitzel, L., Prati, R. C., & Aguiar, R. F. (2016). The comprehension of figurative language: What is the influence of irony and sarcasm on NLP techniques? *Sentiment Analysis and Ontology Engineering: An Environment of Computational Intelligence*, 49-74.

Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual Review of Psychology, 69*, 105-129.

Wies, S., Bleier, A., & Edeling, A. (2022). Finding Goldilocks Influencers: How follower count drives social media engagement. *Journal of Marketing*, *87*(3), 383-405.

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision, 4*(12), 11-11.

Wilson, T. D., & Gilbert, D. T. (2008). Explaining away: A model of affective adaptation. *Perspectives on Psychological Science*, *3*(5), 370-386.

Wojcieszak, M., Casas, A., Yu, X., Nagler, J., & Tucker, J. A. (2021). Echo chambers revisited: The (overwhelming) sharing of in-group politicians, pundits and media on Twitter. *OSF Preprint.* https://doi.org/10.31219/osf.io/xwc79

Wojcieszak, M., Casas, A., Yu, X., Nagler, J., & Tucker, J. A. (2022). Most users do not follow political elites on Twitter; those who do show overwhelming preferences for ideological congruity. *Science Advances, 8*(39), eabn9418.

Wojcik, S., & Hughs, A. (2019). How twitter users compare to the general public. *Pew Research Center*. https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/

Wolfsfeld, G., Segev, E., & Sheafer, T. (2013). Social media and the Arab Spring: Politics comes first. *The International Journal of Press/Politics, 18*(2), 115-137.

Wood-Doughty, Z., Smith, M., Broniatowski, D., & Dredze, M. (2017, August). How does Twitter user behavior vary across demographic groups?. In *Proceedings of the Second Workshop on NLP and Computational Social Science* (pp. 83-89).

Wood, S. (2015). Package 'mgcv'. *R package version, 1*(29), 729.

Yang, C.C. (2016). Instagram use, loneliness, and social comparison orientation: Interact and browse on social media, but don't compare. *Cyberpsychology, Behavior, and Social Networking, 19*(12), 703-708.

Yim, S. H., Lorenz, H., & Salkovskis, P. (2023). The effectiveness and feasibility of psychological interventions for populations under ongoing threat: A systematic review. *Trauma, Violence, & Abuse,*
0(0). https://doi.org/10.1177/15248380231156198

Zaki, J., & Williams, W. C. (2013). Interpersonal emotion regulation. *Emotion, 13*(5), 803.

Zhang, H., & Qu, C. (2018). Emotional, especially negative microblogs are more popular on the web: evidence from an fMRI study. *Brain Imaging Behavior Research Methods*, 1-11.

Zipf, G. K. (1932). *Selected studies of the principle of relative frequency in language.* Cambridge, MA: Harvard University Press.

Zipf, G. K. (1949). *Human behaviour and the principle of least-effort.* Cambridge MA edn. Reading: Addison-Wesley, 24.

APPENDICES

CHAPTER 2: SUPPLEMENTARY MATERIALS

Analysis Using the Full Sample

The goal of this analysis was to replicate the findings reported in the main manuscript using the 2,674,418 tweets of all users, including 2,246,068 tweets produced by 39,241 public figures and 428,350 tweets from 6,677 ordinary users. I repeated every analysis as reported in the main manuscript using this full sample, except for the *Association between User Characteristics and Negativity Sharing* section. Differences in analysis were described in the corresponding sections.

Association between Sentiment and Retweets for Public Figures versus Ordinary Users.

For this analysis, the dependent variable was content sharing, measured by the number of retweets (log+1-transformed). For the independent variable, we used the continuous sentiment score from VADER (-1 extremely negative to +1 extremely positive). I fitted a quadratic mixed model to predict the number of retweets using a quadratic function of the continuous sentiment score between -1 and 1.

Table 2.1S. *Quadratic mixed model with five factors (Sentiment^2, Sentiment^1, User_type, Sentiment^1 * User_type, Sentiment^2 * User_type) and number of retweets (log+1) as the dependent variable.*

Fixed Effects					
	Estimate	*SE*	95% CI	*t*	*p*
Intercept	0.20	0.11	0.18 – 0.22	18.63	.000
Sentiment^1	0.0048	0.002	-0.00095 – 0.010	1.63	.10
Sentiment^2	0.048	0.0055	0.037 – 0.059	8.78	.000
User_type:Public Figure	0.61	0.011	0.59 – 0.63	51.55	.000

Sentiment[1] × User_type:Public Figure	-0.10	0.0032	-0.10 – -0.095	-31.23	.000
Sentiment[2] × User_type:Public Figure	0.15	0.0060	0.14 – 0.16	25.90	.000

Random Effects

	Variance	*SD*
User_id (Intercept)	0.82	0.90
Residual	0.64	0.80

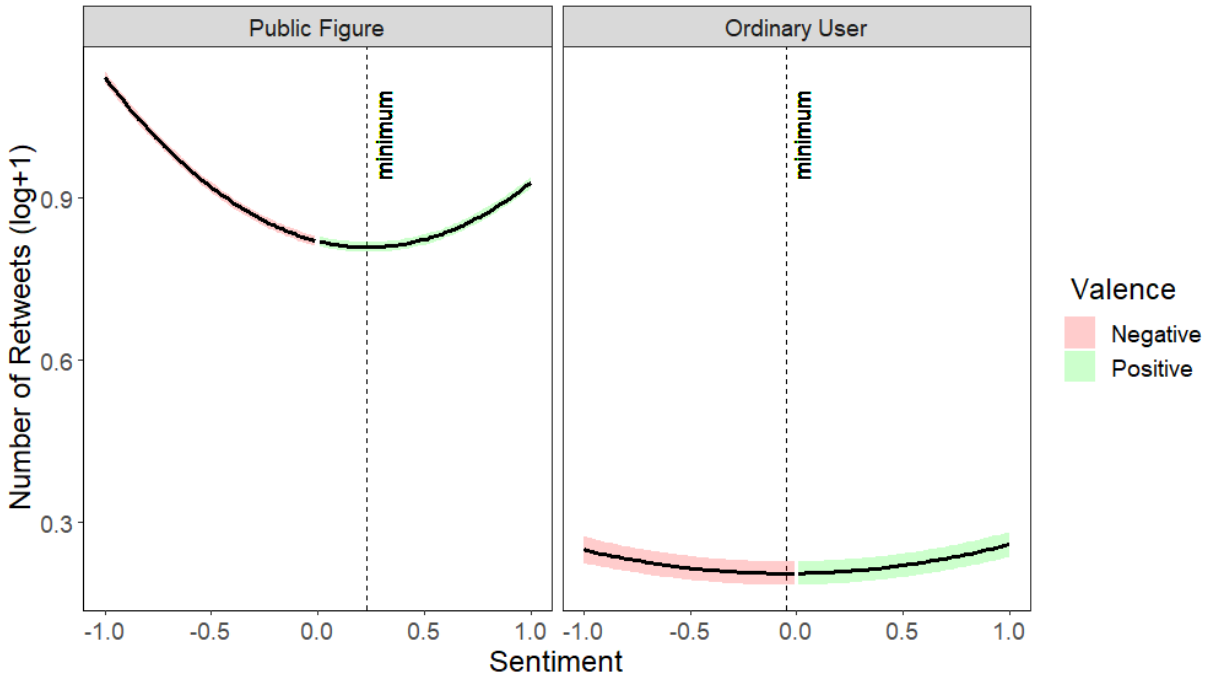
Model Fit

R^2	Marginal	Conditional
	0.037	0.57

Model equation: log(retweets+1) ~ poly(compound,2, raw = TRUE) * user_type + (1 | User)

Notes. Model fit was calculated using the R package MuMIn (Barton & Barton, 2015) based on the paper of Nakagawa et al. (2017).

Figure 2.1S. *Number of retweets as a function of sentiment and user type.*



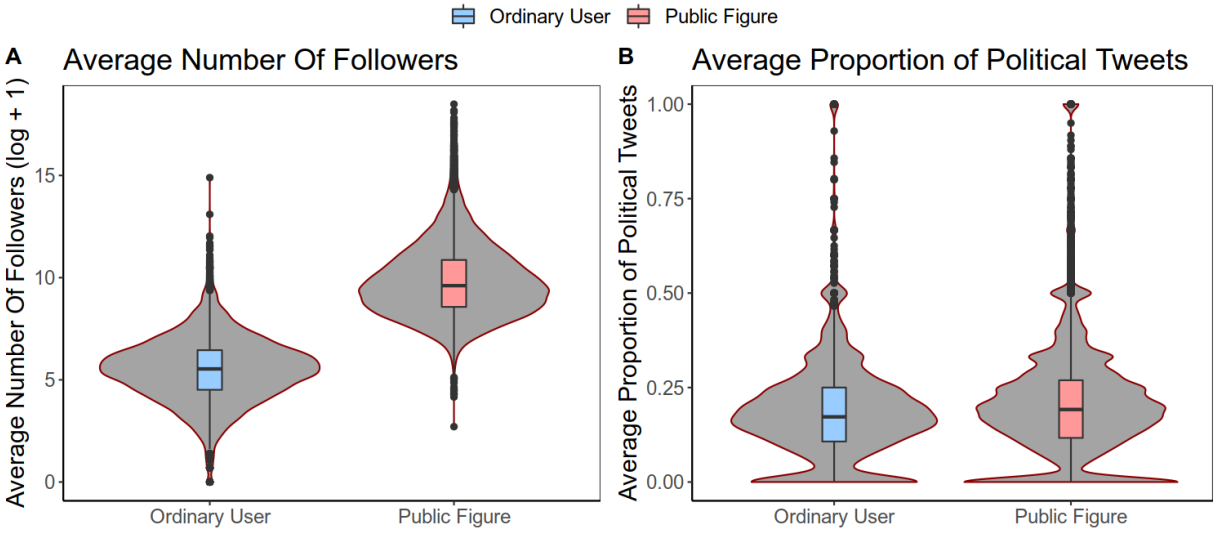
Notes. *The results suggest that stronger sentiment is associated with more retweets for both types of users. The local minimum for public figures is reached with a more positive emotional tweet, indicating that negativity is more strongly positively associated with number of retweets for public figures than for ordinary users. Public figures also receive more retweets for neutral content than ordinary users.*

Differences between Public Figures and Ordinary Users.

I tested if public figures had more followers, and a higher proportion of political tweets than ordinary users, using the aggregated user characteristics measures. These were the log+1 transformed number of followers of a user, and the proportion of political tweets defined as the number of tweets identified as political tweets by the topics modelling divided by the number of total tweets. I used a simple linear regression predicting the two user characteristics with a dummy coded variable for verification status.

As reported in the main manuscript, the analysis using the full dataset replicated the finding in the main manuscript that public figures had more followers (b = 4.33 [4.28, 4.37], SE = 0.022, t (45911) = 190.2, p < .001, R^2 = .44, see Figure 2.2S.A), and produced content that contained approximately 2% more political content (b = 0.019 [0.015, 0.025], SE = 0.0020, t (45911) = 9.80, p < .001, R^2 = .0020, see Figure 2.2S.B) than ordinary users.

Figure 2.2S. *Differences between public figures and ordinary users.*



Notes. *Results suggest that public figures have more followers than ordinary users (Panel A), and produce a higher proportion of political tweets (Panel B).*

Association between User Characteristics and Negativity Sharing.

I tested if user characteristics interacted with the relationship between sentiment and the number of retweets. I again used a quadratic mixed model to predict the number of

retweets using a quadratic function of the continuous sentiment score. I included each user characteristic as an interaction factor with sentiment in two models separately and added a random intercept for users to account for baseline differences in the average number of retweets that they get. In contrast to the analysis reported in the main manuscript, we added a covariate for the different user types into the models to account for the differences in sample size, as the current sample contains more verified users than ordinary users. This should avoid that the effect of user characteristics is driven by the more frequent verified users.

Number of Followers

Table 2.2S. *Quadratic mixed model with six factors (Sentiment^2, Sentiment^1, Number_of_Followers, User_type, Sentiment^1 * Number_of_Followers, Sentiment^2 * Number_of_Followers) and number of retweets (log+1) as the dependent variable.*

Fixed Effects					
	Estimate	*SE*	95% CI	*t*	*p*
Intercept	1.31	0.012	1.28 – 1.33	101.70	.000
Sentiment^1	-0.081	0.0012	-0.083 – -0.078	-65.10	.000
Sentiment^2	0.17	0.0022	0.17 – 0.18	79.39	.000
Number_of_Followers	0.66	0.0051	0.65 – 0.67	129.38	.000
User_type:Public Figure	-0.45	0.013	-0.47 – -0.42	-33.98	.000
Sentiment^1 × Number_of_Followers	-0.038	0.0012	-0.041 – -0.036	-31.97	.000
Sentiment^2 × Number_of_Followers	0.047	0.0022	0.043 – 0.052	21.41	.000

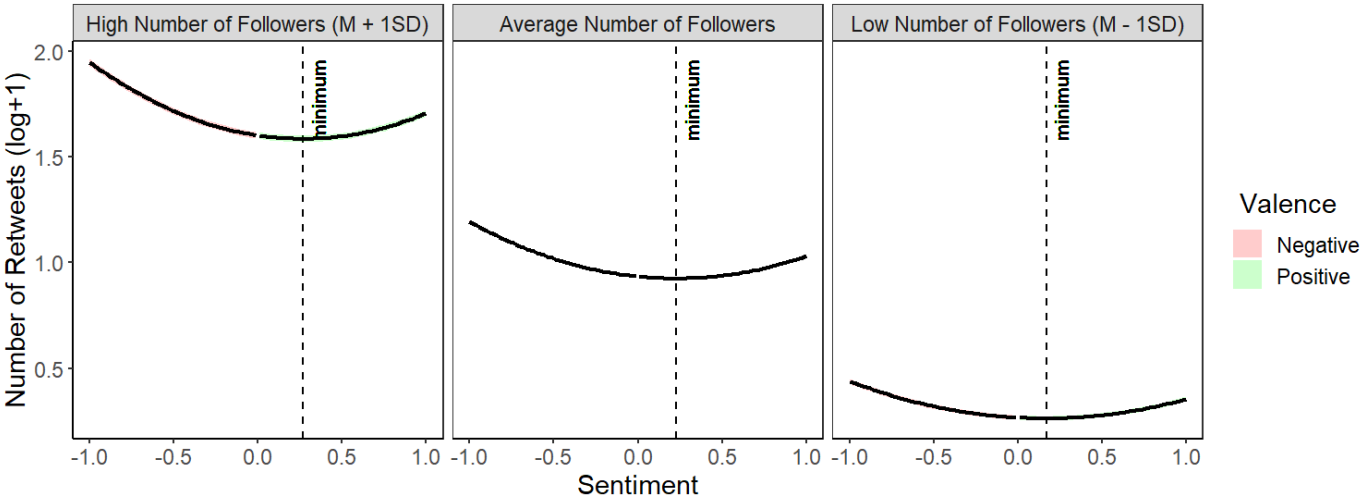
Random Effects		
	Variance	*SD*
User_id (Intercept)	0.58	0.76
Residual	0.64	0.80

Model Fit		
R^2	Marginal	Conditional
	0.21	0.58

Model equation: log(retweets+1) ~ poly(compound,2, raw = TRUE) * scaled(Number_of_Followers) + (1 | User)

Notes. *Model fit was calculated using the R package MuMIn (Barton & Barton, 2015) based on the paper of Nakagawa et al. (2017).*

Figure 2.3S. *The number of retweets as a function of sentiment and number of followers.*



Notes. *To visualize the interaction of two continuous variables (sentiment and the number of followers), the panels show the mean number of followers in the middle panel as well as the mean plus and minus one standard deviation (left and right panel respectively). The results showed that an increase in the number of followers strengthens the link between sentiment and content sharing. Negativity sharing was stronger for the content of users with more followers.*

Proportion of Political Tweets

Table 2.3S. *Quadratic mixed model with six factors (Sentiment^2, Sentiment^1, Political_Tweets, User_type, Sentiment^1 * Political_Tweets, Sentiment^2 * Political_Tweets) and number of retweets (log+1) as the dependent variable.*

Fixed Effects					
	Estimate	*SE*	95% CI	*t*	*p*
Intercept	0.19	0.011	0.17 – 0.21	17.55	.000
Sentiment^1	-0.077	0.0012	-0.080 – -0.075	-61.93	.000
Sentiment^2	0.17	0.0022	0.17 – 0.18	77.31	.000
Political_Tweets	0.021	0.0027	0.015 – 0.026	7.73	.000

User_type:Public Figure	0.63	0.011	0.60 – 0.65	53.17	.000
Sentiment[1] × Political_Tweets	-0.013	0.0013	-0.015 – -0.010	-10.08	.000
Sentiment[2] × Political_Tweets	0.0059	0.0023	0.0013 – 0.010	2.54	.000

Random Effects

	Variance	*SD*
User_id (Intercept)	0.82	0.90
Residual	0.64	0.80

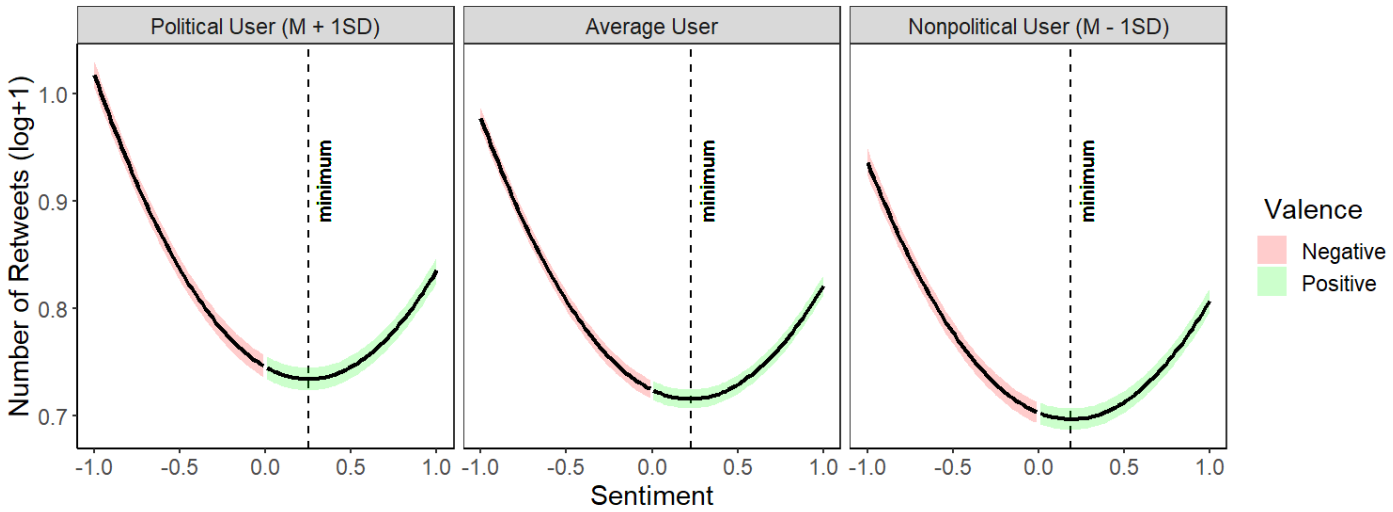
Model Fit

R^2	Marginal	Conditional
	0.037	0.58

Model equation: log(retweets+1) ~ poly(compound,2, raw = TRUE) * scaled(Political_Tweets) + (1 | User)

Notes. *Model fit was calculated using the R package MuMIn (Barton & Barton, 2015) based on the paper of Nakagawa et al. (2017).*

Figure 2.4S. *The number of retweets as a function of sentiment and proportion of political tweets.*



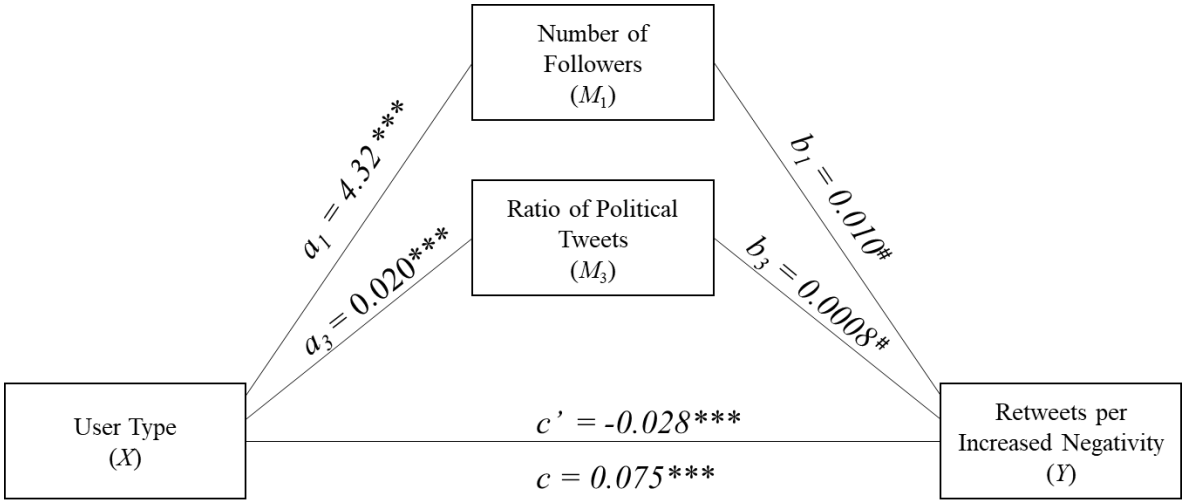
Notes. *To visualize the interaction of two continuous variables (sentiment and proportion of political tweets), the panels show the predicted association between sentiment and the number of retweets at the mean proportion of political tweets in the middle panel as well as the mean plus and minus one standard deviation (left and right panel respectively). The results indicate that tweets produced by users with a higher proportion of political tweets show stronger associations between sentiment and content sharing and between negativity and content sharing.*

Parallel Mediation Analysis.

Finally, we repeated the parallel mediation analysis assessing two potential mediators of the effects of user type (public figures vs. ordinary users) on sharing of their negative content (Hayes, 2017) as described in the manuscript. The potential mediators were the two user characteristics, i.e., number of followers, and the proportion of political tweets. I again calculated the dependent variable that quantified the extent to which an increase in negativity was associated with more retweets for every individual user in our dataset (see manuscript for all details). I used the PROCESS v4 macro for RStudio by Hayes (2017) to conduct the parallel mediation analysis.

As reported above, user type correlated positively with number of followers (a_1 = 4.32 [4.28, 4.37], SE = 0.022, *t* (45913) = 190.032, *p* <.001, R^2 = 0.09, see Figure 2.5S), and the proportion of political tweets (a_2 = 0.020 [0.016, 0.024], SE = 0.0020, *t* (45913) = 6.56, *p* < .001, R^2 = 0.09). A 95% bias-corrected confidence interval based on 10,000 bootstrap samples indicated that the sampled indirect effects through number of followers (a_1b_1 = 0.10, SE = 0.0024), holding all other mediators constant, were consistently above zero (0.098 - 0.10). The indirect effects through proportion of political tweets were consistently above zero (a_3b_3 = 0.0006 [0.0001 - 0.0008], SE = 0.0010). The direct total affect in the absence of the mediator yielded a significant association between verification status and negativity sharing (c = 0.075 [0.071, 0.079], SE = 0.0021, *t* (45913) = 36.01, *p* <.001, R^2 = 0.09). In contrast to the findings in the main manuscript, the total effect of user type on negativity sharing was also significant after controlling for the mediators (c' = -0.028 [-0.034 - -0.023], SE = 0.0027, *t* (45913) = -10.67, , *p* <.001, R^2 = 0.09). These results from the parallel mediation analysis suggest that user type only had an effect on negativity sharing, but was partially mediated by the number of followers and proportion of political tweets.

Figure 2.5S. *Parallel mediation analyses the effect of user type on negativity sharing via three different user characteristics*.



Notes. *User type positively predicts each of the user characteristics: number of followers, and proportion of political tweets. The indirect effects of user type on negativity sharing via the number of followers and proportion of political tweets were both significant.*

Sentiment Tool Comparison

To test whether VADER is the best sentiment analysis tool for the dataset, we first tested the correlation between VADER and another commonly used sentiment analysis tool, SentiStrength (Pennebaker et al., 2001). SentiStrength is comprised of two independent sentiment scores, one for positive language and the other for negative language. I computed a total sentiment score by subtracting a tweets negative sentiment score from the positive sentiment score. The correlation of the total sentiment scores of VADER and SentiStrength was r = 0.56 (95%CI = 0.55 - 0.57, t(855713)= 633.99, p < .001).

I repeated the analysis testing whether negativity is more positively associated with shares for public figures than for ordinary users by using the identical model as for the VADER analysis. Namely, we predicted the log-modulus transformed number of retweets by the SentiStrength sentiment score and its interaction with user type using a quadratic mixed model with a random intercept of individual users.

Table 2.4S. *Quadratic mixed model with five factors (SentiStrength_Sentiment^2, SentiStrength_Sentiment^1, User_type, SentiStrength_Sentiment^1 * User_type, SentiStrength_Sentiment^2 * User_type) and number of retweets (log+1) as the dependent variable.*

Fixed Effects

	Estimate	*SE*	95% CI	*t*	*p*
Intercept	0.20	0.0090	0.18 – 0.22	22.4	.000
SentiStrength_Sentiment[1]	-0.0035	0.00080	-0.0051 – -0.0019	-4.43	.000
SentiStrength_Sentiment[2]	-0.00015	0.00041	-0.00097 – 0.00065	-0.37	.71
User_type:Public Figure	0.61	0.012	0.58 – 0.63	49.53	.000
SentiStrength_Sentiment[1] × User_type:Public Figure	-0.036	0.0011	-0.039 – -0.034	-31.36	.000
SentiStrength_Sentiment[2] × User_type:Public Figure	0.0025	0.00059	0.0014 – 0.0037	4.34	.000

Random Effects

	Variance	*SD*
User_id (Intercept)	0.53	0.73
Residual	0.46	0.68

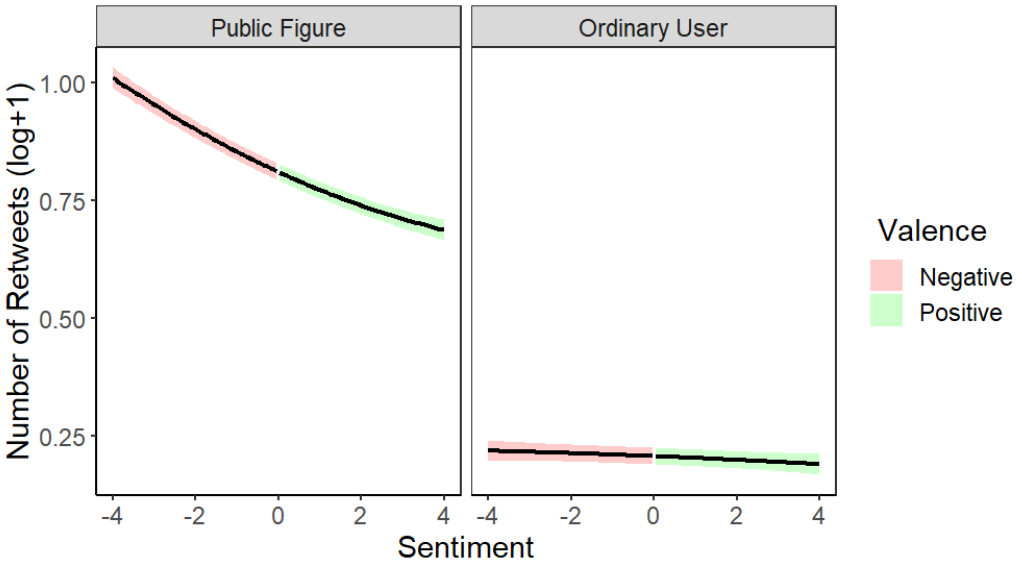
Model Fit

R^2	Marginal	Conditional
	0.087	0.57

Model equation: log(retweets+1) ~ poly(SentiStrength, raw = TRUE) * user_type + (1 | User)

Notes. Model fit was calculated using the R package MuMIn (Barton & Barton, 2015) based on the paper of Nakagawa et al. (2017).

Figure 2.6S. *The number of retweets as a function of sentiment as measured by SentiStrength.*



Notes. *The results showed, more negativity is associated with an increase in the number of retweets. This effect was stronger for public figures than for ordinary users, replicating the main finding of the manuscript.*

Finally, we compared the model fit between the models using VADER compared to SentiStrength. The model using VADER had the overall better model fit (AIC = 1,818,927) in comparison to SentiStrength (AIC = 1,819,941), we therefore decided to use VADER for all other analysis.

Topics Modelling Detailed Information

Political Topic Presented in the Manuscript.

This section is meant to provide more details regarding the procedure applied to determine the political topic used to classify political tweets. I analysed the meaning of each of the 5 topics to find the one that is about politics. In order to understand the meaning of a topic, we first analysed the overarching theme of the words that have the highest probability to be generated by a certain topic, which is expressed by the β-score. After manually inspecting the 5 topics, we identified one topic that seemed to generate politics related word with a higher likelihood (Table 2.5S). A full list of the most common words per topics is available on the OSF repository (topic_modell_words_5t.csv). I confirmed the quality of the topic classification by inspecting the tweets that had the highest likelihood of containing this topic as indicated by the γ-score (Table 2.6S).

Table 2.5S. *The 20 most frequent unigrams for the political topic.*

Political Topic	β-score
Trump	0.0079
time	0.0056
day	0.0052
feel	0.0047
talk	0.0045
vote	0.0044
week	0.0041
happen	0.0040
share	0.0038
win	0.0038
life	0.0037
shut-down	0.0037
people	0.0036
live	0.0036
president	0.0035
update	0.0032
city	0.0031
learn	0.0030
shutdown	0.0030

Notes. This table shows the words showing the biggest β, which indicates the frequency of a word in a given topic. I concluded that this topic based on the words most frequent in this topic is related to politics

Table 2.6S. Most political tweets: *Tweets with the highest γ-score for the above mentioned political topic, meaning that these tweets should contain political content with the highest likelihood according to our topic modelling analysis.*

No.	Tweet content	γ-score
1.	McGurk decry 'total reversal' of # us policy on # syria	0.56
2.	Sexual harassment and pay disparity in his campaign complicate Bernie Sander. The former presidential candidate's desire to return to the ring is complicated by several testimonies revealed by 'The New York Times'	0.56

3.	the man accuse in the fatal shoot of lsu basketball player Wayde Sims have be indict on a murder charge	0.47
4.	API investor Andy Gracey say a proposal from the drug wholesaler to merge with sigma healthcare throw a lifeline to a under - pressure rival, @ jessigardner report.	0.44
5.	hook, line, and sinker - what do you know about # phishing? it's a effective and often successful practice, so how do you prevent it? nleefvejc # cybersecurity	0.43
6.	turkey: pres. erdogan blast u. s. change in syria withdrawal plan	0.25
7.	Opinion: the shutdown, which change the ceremony, have also deeply affect a group # mlk advocate for - people experience poverty.	0.24
8.	The partial government shutdown, now the under president donald trump, be the long shutdown in u. s. history as of today (jan.). its impact be far reach in # washington, strain earthquake research and, soon, food assistance.	0.24
9.	There's plenty of speculation over the reason u. s. president Donald Trump have call a prime - time national address, many believe it's a new attempt to end the stalemate over his mexican border wall.	0.24
10.	Probe governor Obado over new murder scheme	0.24

Political Topic Assessed Using Topics Modelling k = 50 and PCA

Creating Alternative Classification

The number of topics in topics modelling is selected manually. As mentioned in the manuscript, we tried to identify the minimum number of topics that contained mostly political words in order to avoid having multiple politics topics. As an alternative test, instead of using the smallest number of topics that includes a political topic, we used larger numbers of topics (k = 50), potentially including multiple but more specific political topics. The manual assessment of the topic meanings confirmed that there were multiple topics related to politics (e.g. see Table 2.7S for the three topics which are associated with the final politics variable as determined by the following procedure, see topic_modell_words_50t.csv for full list of words in topics). However, in order to conduct the analysis using a single politics classification variable, we used Principle Component Analysis (PCA) to reduce the dimensionality of the 50 topics by clustering those share the most variance with the goal to identify one component that is comprised of political topics. I decided to conduct a PCA analysis with 5 components to match the number of components in the topic modelling presented in the manuscript. To identify the component that represents political content, we analysed the semantic meaning of the topics that had the highest factor loadings for each component. The political component

was therefore the one that had the highest factor loadings from various political topics (Table 8S). I then classified a tweet as political if the component containing the political topics had the factor score compared to the other four components.

Table 2.7S. *The 5 most frequent unigrams for the political topics that contribute most to the political component.*

Topic	Most Frequent Word	β-score
A	Trump	0.061
A	wall	0.052
A	president	0.039
A	border	0.032
A	news	0.027
B	shutdown	0.018
B	government	0.010
B	question	0.0089
B	Trump	0.0081
B	fund	0.0080
C	government	0.023
C	people	0.018
C	time	0.015
C	color	0.013
C	risk	0.012

Notes. *This table shows the words showing the biggest β, which indicates the frequency of a word in a given topic. I concluded that this topic based on the words most frequent in this topic is related to politics*

Table 2.8S. *Factor contribution of the political topics to the political factor*

Political Component	Topic Contribution
Topic A	16.94
Topic B	8.74
Topic C	7.65

Notes. *For each political topic described in the previous table, this table shows how much they contribute to the political component that is used for classification (in %).*

Comparing Alternative Classification to Model presented in the Manuscript

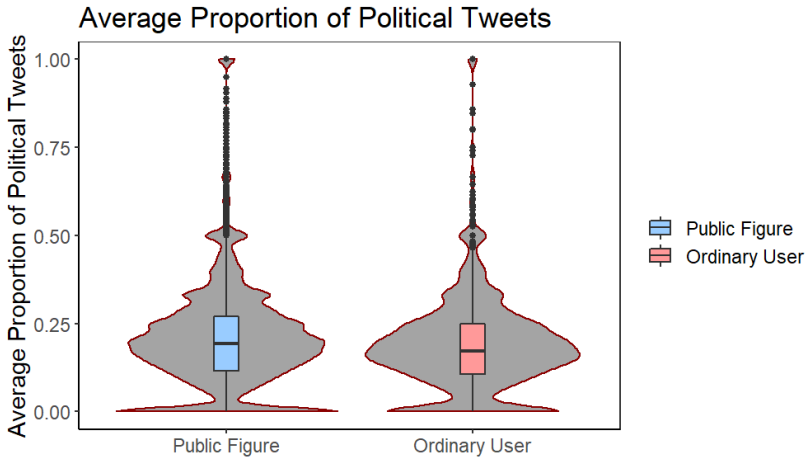
I tested how correlated the classification of political tweets was using the method described in the manuscript (topic modelling, k = 5) with the above-mention method (topic modelling, k = 50, PCA =5). The correlation between the estimates was r = 0.14 (95%CI = 0.55 - 0.57, t(857740)= 138.7, p < .001). I also tested how correlated the estimation of the user characteristics, proportion of political tweets (calculated as the number of political tweets compared to all tweets), was using both tools correlated with r = 0.22 (95%CI = 0.21 - 0.23, t(45911)= 48.28, p < .001).

Repeating Analysis Using Alternative Classification

I repeated the analysis regarding the proportion of political tweets. First, we compared the difference in the proportion of political tweets of public figures compared to ordinary users. Before also testing how the proportion of political tweets influences the likelihood of negativity being shared for someone.

In line with the findings presented in the main manuscript, public figures produced more political content (b = 0.04 [0.037, 0.046], SE = 0.0022, t (45911) = 18.86, p < .001, R^2 = .0076, see Figure 7S) than ordinary users.

Figure 2.7S. *Differences between public figures and ordinary users.*



Notes. *Results suggest that public figures produce around 4% more political content than ordinary users.*

I again used a quadratic mixed model to predict the number of retweets using a quadratic function of the continuous sentiment score. I included the alternative estimate for proportion of political content (topic model, k = 50, PCA = 5) as an interaction factor with sentiment in the model and added a random intercept for users to account for baseline differences in the average number of retweets that they get.

Table 2.9S. *Quadratic mixed model with five factors (Sentiment^2, Sentiment^1, Political_Tweets, Sentiment^1 * Political_Tweets, Sentiment^2 * Political_Tweets) and number of retweets (log+1) as the dependent variable.*

Fixed Effects					
	Estimate	*SE*	95% CI	*t*	*p*
Intercept	0.50	0.0073	0.48 – 0.51	68.79	.000
Sentiment^1	-0.048	0.0018	-0.051 – -0.044	-25.94	.000
Sentiment^2	0.13	0.0034	0.12 – 0.13	38.16	.000
Political_Tweets	0.093	0.0056	0.082 – 0.10	16.73	.000
Sentiment^1 × Political_Tweets	-0.020	0.0017	-0.023 – -0.016	-11.75	.000
Sentiment^2 × Political_Tweets	0.011	0.0032	0.0044 – 0.017	3.38	.000

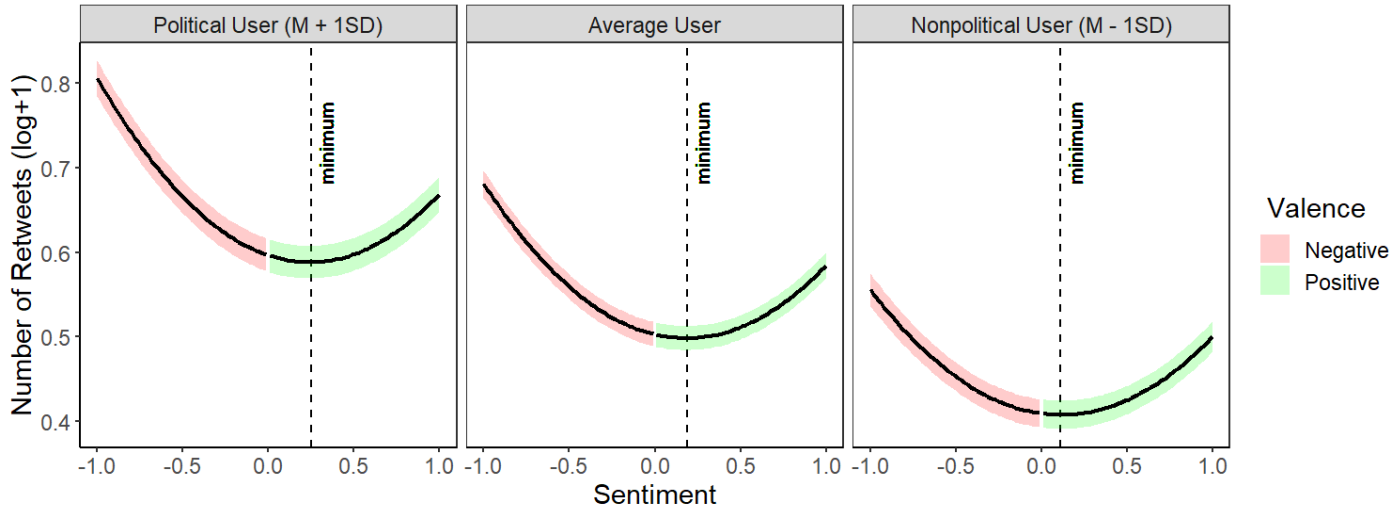
Random Effects		
	Variance	*SD*
User_id (Intercept)	0.82	0.90
Residual	0.64	0.80

Model Fit		
R^2	Marginal	Conditional
	0.0097	0.58

Model equation: log(retweets+1) ~ poly(compound,2, raw = TRUE) * scaled(Political_Tweets) + (1 | User)

Notes. Model fit was calculated using the R package MuMIn (Barton & Barton, 2015) based on the paper of Nakagawa et al. (2017).

Figure 2.8S. *The number of retweets as a function of sentiment and proportion of political tweets.*



Notes. *To visualize the interaction of two continuous variables (sentiment and proportion of political tweets), the panels show the predicted association between sentiment and the number of retweets at the mean proportion of political tweets in the middle panel as well as the mean plus and minus one standard deviation (left and right panel respectively). The results indicate that tweets produced by users with a higher proportion of political tweets show stronger associations between sentiment and content sharing and between negativity and content sharing.*

Analysing Different Types of Verified Users

On Twitter, various types of public figures can obtain verification status, including politicians, journalists, celebrities to athletes. Some of verified users are individuals while others represent organizations such as companies, universities or government institutions. This raises the question whether the main finding, namely, that negative content is shared more for verified users, hold true for all types of verified users. Nonetheless, we hypothesize that the tendency for negative content to be shared more among verified users is due to their common characteristics, such as having a large number of followers and discussing political topics more often. Therefore, we still anticipate that negativity will be shared more frequently among the majority of verified user types, especially news media and political figures who are more likely to engage in political discourse.

I tried to identify several major categories of verified users, including political figures, journalists, news organizations, entertainment, sports and organizations, and evaluated the

tendency of their negative content to be shared by other users. To do so, employed three classification approaches in conjunction to evaluate these categories. First, we used manually curated dictionaries on the profile descriptions, which were informed by our topic modelling that provided me with a selection of frequent words that might make up one of these categories. For example, to identify political figures, we used partial matching to search for specific word parts such as "congress," "parliament," "senator," "governor," "mayor," and "council" (see Table 2.10S for all categories and corresponding word parts). Second, we leveraged to previous research on verified users, such as the lists created by Barberá (2015) as well as Rathje et al. (2021) for political figures, and the documentation by Bellovary et al. (2021) for media outlets. I then linked the users identified as certain public figure types in these lists with our sample. Finally, we employed the tool "Demographer" which utilizes machine learning and natural language processing techniques to infer whether an account belongs to an individual or organization from multilingual social media data. Using these techniques, we were able to categorize around 46.96% of all verified users into accounting for 55.45% of the tweets in our dataset.

Table 2.10S. *Word parts for each dictionary category.*

Public Figure Category	Dictionary Word Parts
Entertainment	actor, actress, producer, writer, comedian, artist, music
Journalists	media, news, journal, reporter
Organisation	company, corporation, enterprise, group
Politics	congress, parliament, senator, governor, mayor, council, commissioner, activist
Sport	athlete, sport, player, referee

Notes. *The table summarizes the word parts used for partial matching of public figures categories in the first step of categorization.*

Average Emotion Expressed by Users

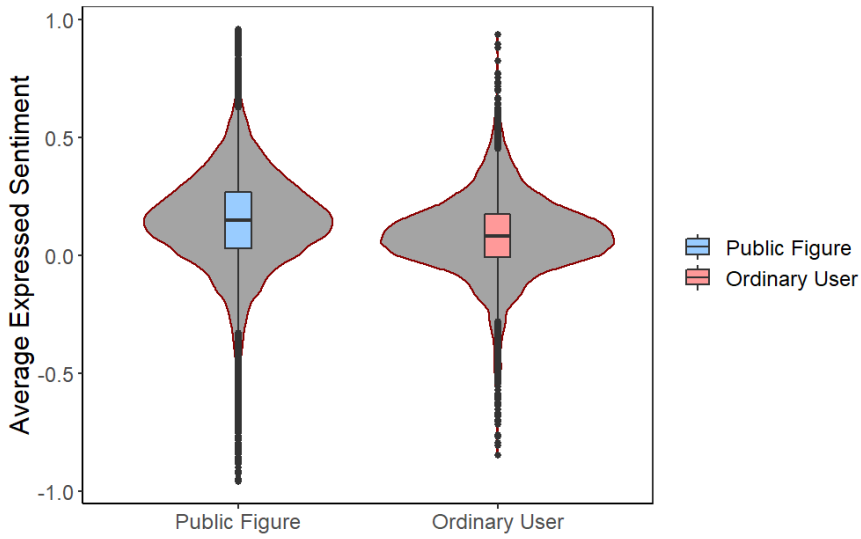
As outlined in the main manuscript, one notable difference between public figures and ordinary users was the prevalence of *positive language* in their posts. This finding may be attributed to the fact that social media is often used by individuals, including public figures, as a tool to enhance their public image (Citra, 2018). The larger a user's network is, the more concerned they are about the impression they make (Hogan, 2010), which motivates them to express themselves more positively to avoid being perceived negatively (Buck, Losow,

Murphy, & Costanzo, 1992; Lin et al., 2014). Public figures with their exceptionally large followings are therefore prone to generating more content with a positive sentiment. For users like public figures who primarily generate positive content, negativity could be shared more, as it is rarer and unexpected.

To test this idea, we calculated the average sentiment expressed by a user as the mean of their individual tweet sentiment. I then first tested the difference in emotion expressed between user types using this average sentiment score instead of using the categorical variables as in the main manuscript, before again testing the influence of these user characteristics on the likelihood that negativity is shared.

Replicating the previous analysis, public figures express themselves more positively (b = 0.064 [0.058, 0.069], SE = 0.0033, t (45911) = 19.00, $p < .001$, $R^2 = .012$, see Figure 2.10S) than ordinary users.

Figure 2.9S. *Differences between public figures and ordinary users.*



Notes. *Results suggest that public figures express themselves using more positive language on average.*

I again used a quadratic mixed model to predict the number of retweets using a quadratic function of the continuous sentiment score. I included the average sentiment expressed by a user as an interaction factor with sentiment in the model and added a random intercept for users to account for baseline differences in the average number of retweets that they get.

Table 2.11S. *Quadratic mixed model with five factors (Sentiment^2, Sentiment^1, Average_Sentiment_Expressed, Sentiment^1 * Average_Sentiment_Expressed, Sentiment^2 * Average_Sentiment_Expressed) and number of retweets (log+1) as the dependent variable.*

Fixed Effects					
	Estimate	*SE*	95% CI	*t*	*p*
Intercept	0.46	0.0086	0.44 – 0.47	53.61	.000
Sentiment^1	-0.034	0.0022	-0.038 – -0.029	-15.50	.000
Sentiment^2	0.16	0.0042	0.15 – 0.16	37.84	.000
Average_Sentiment_Expressed	0.21	0.038	0.13 – 0.29	5.52	.000
Sentiment^1 × Average_Sentiment_Expressed	-0.15	0.0014	-0.17 – -0.11	-10.47	.000
Sentiment^2 × Average_Sentiment_Expressed	-0.081	0.0022	-0.12 – -0.038	-3.69	.000

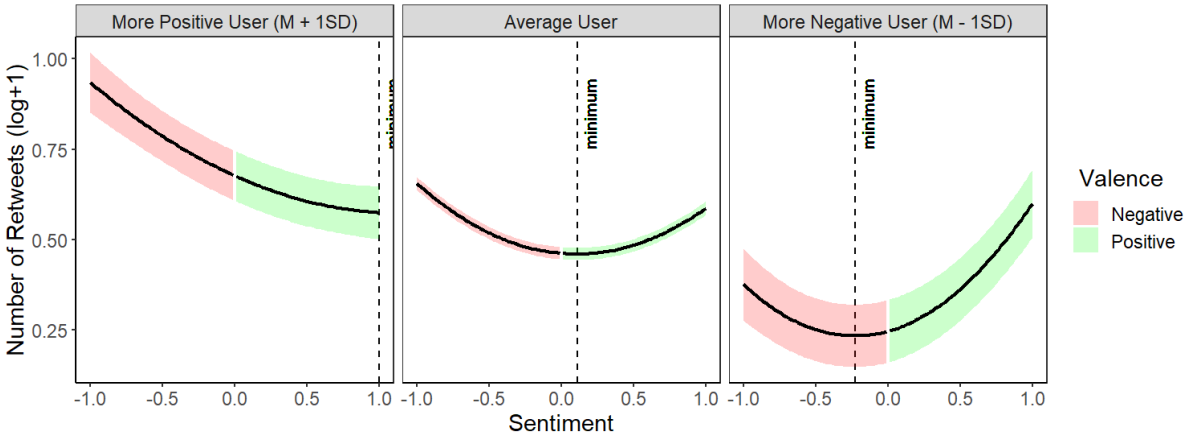
Random Effects		
	Variance	*SD*
User_id (Intercept)	0.66	0.81
Residual	0.46	0.68

Model Fit		
R^2	Marginal	Conditional
	0.0014	0.59

Model equation: log(retweets+1) ~ poly(compound,2, raw = TRUE) * scaled(Average_Sentiment_Expressed) + (1 | User)

Notes. *Model fit was calculated using the R package MuMIn (Barton & Barton, 2015) based on the paper of Nakagawa et al. (2017).*

Figure 2.10S. *The number of retweets as a function of sentiment and average sentiment expressed.*



Notes. *To visualize the interaction of two continuous variables (sentiment and average sentiment expressed), the panels show the predicted association between sentiment and the number of retweets at the mean sentiment expressed in the middle panel as well as the mean plus and minus one standard deviation (left and right panel respectively). The results indicate that tweets produced by users with more positive tweets on average show stronger associations between sentiment and content sharing and between negativity and content sharing.*

General Additive Model of Sentiment and Retweets for Public Figures and Ordinary Users

To verify whether quadratic models were the best choice, we conducted a general additive mixed model analysis using the mgcv package in R (Wood, 2015), on the full sample including 2,674,418 tweets of all users by 39,241 public figures and 6,677 ordinary users. Similar to the quadratic mixed model, we used the continuous VADER sentiment score to predict the log-modulus transformed retweet count. I chose four knots for the sentiment score, which determine the number of splines used for the fitting function. However, since general additive models do not allow for interactions of the splines, we conducted separate analyses for each user type. As use the full sample, that is not matched by their online activity, we included total tweets produced as a parametric covariate. Our findings confirmed the earlier observation that negativity spreads more widely than positivity. Furthermore, the results demonstrated that the quadratic function provided a good approximation of the relationship between sentiment and sharing.

GAM for Public Figures.

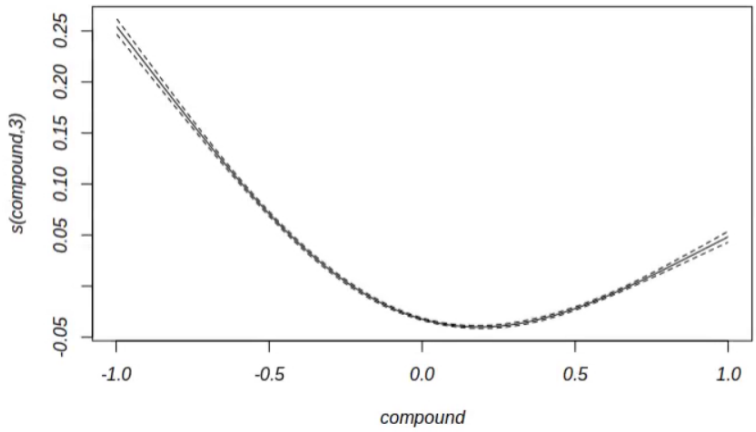
Table 2.12S. GAM model: *General additive model with one smooth factors (sentiment score) using 4 knots and one parametric parameter (total number of tweets) predicting number of retweets as the dependent variable.*

Parametric Terms				
	Estimate	*SE*	*t*	*p*
Intercept	-0.098	0.0018	-53.38	.000
Total Tweets	0.039	0.00031	124.61	.000
Smooth-Terms				
	edf	Ref.df	*F*	*p*
Sentiment (4 knots)	1.00	1.00	1.25	.884
Random Effects				
		Variance	*SD*	
Participant (Intercept)		0.017	0.13	
Residual		0.034	0.18	
Model fit				
R^2			Adjusted	
			.14	

Model equation: log(retweets+1) ~ (compound, knots = 4) + total_tweets + (1 | User)

Notes. *General additive mixed model was computed via mgcv package in R (Wood, 2015).*

Figure 2.11S. *Results of sentiment predicting number of retweets for the content produced by public figures.*



Notes. *I was able to replicate the previous finding, that negativity is shared more for public figures using GAM, which allows the function to be fitted more flexible to the data (4 knots). After visual inspection, we concluded that a quadratic function is indeed an appropriate approximation of the relationship between sentiment and retweets.*

GAM for Ordinary Users

Table 2.13S. GAM model: *General additive model with one smooth factors (sentiment score) using 4 knots and one parametric parameter (total number of tweets) predicting number of retweets as the dependent variable.*

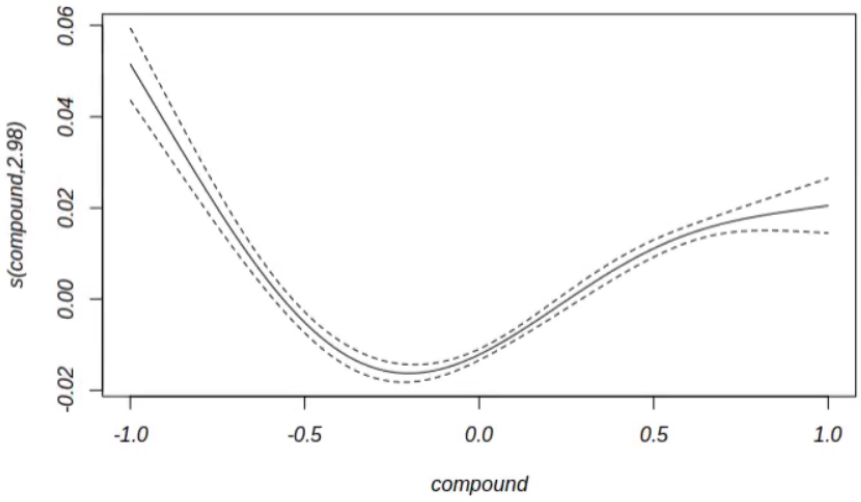
Parametric Terms				
	Estimate	*SE*	*t*	*p*
Intercept	0.15	0.0047	32.52	.000
Total Tweets	-0.00018	0.000036	-5.20	.000
Smooth-Terms				
	edf	Ref.df	*F*	*p*
Sentiment (4 knots)	2.98	2.98	133.2	.002
Random Effects				

	Variance	*SD*
Participant (Intercept)	0.092	0.30
Residual	0.17	0.41
Model fit		
R^2		Adjusted
		.09

Model equation: log(retweets+1) ~ (compound, knots = 4) + total_tweets + (1 | User)

Notes. *General additive mixed model was computed via mgcv package in R (Wood, 2015).*

Figure 2.12S. *Results of sentiment predicting number of retweets for the content produced by Ordinary Users.*



Notes. *I was able to replicate the previous finding, that both positivity and negativity is associated with increased shares using GAM, which allows the function to be fitted more flexible to the data (4 knots). After visual inspection, we concluded that a quadratic function is indeed an appropriate approximation of the relationship between sentiment and retweets.*

CHAPTER 3: SUPPLEMENTARY MATERIALS

Emotional Language Intensity Predicting Likes in Studies 2 and 3

I assessed the degree to which positive and negative emotion scores predicted the spread of content as measured by the number of likes. For both studies, we conducted linear mixed model analysis predicting the number of likes by both the positive emotion score and the negative emotion score for each tweet as well as their interaction. As many tweets did not receive any retweets, the distribution of retweets was positively skewered around zero. I therefore did a reciprocal transformation denoted by $f(x) = 1-(x+1)^{-1}$. A reciprocal transformation involves dividing 1 by the retweet number. Since many of our between numbers were zeros, we first conducted a x+1 transformation. Finally, while reciprocal transformation reduces the skewness in the data it also reveres the numbers' magnitude. To account for this, we subtracted the number from 1, which was the largest number after the transformation. It is noteworthy that the distribution is still not strictly normally distributed due to the excess zeros. I again added a random intercept of user id to the model to deal with the issue of multiple tweets from the same users. I also included the user's number of followers as a covariate, because users with more followers generally have more likes regardless of the emotional content of their tweets.

Hillary election loss

Table 3.1S. *Linear mixed model with four factors (positive language, negative language, number of followers, positive language x negative language) and number of likes 1-(reciprocal+1) as the dependent variable.*

Fixed Effects					
	Estimate	*SE*	95% CI	*t*	*p*
Intercept	0.29	0.00088	0.29 – 0.30	337.33	.000
Positive Language	0.0093	0.00090	0.0075 – 0.011	10.32	.000
Negative Language	0.0124	0.00050	0.011 – 0.013	24.45	.000
Number of followers	0.077	0.00068	0.076 – 0.079	112.69	.000
Positive Language × Negative Language	-0.00088	0.00060	-0.0020 – 0.00020	-1.46	.144
Random Effects					

	Variance	*SD*
Participant (Intercept)	0.028	0.17
Residual	0.057	0.24

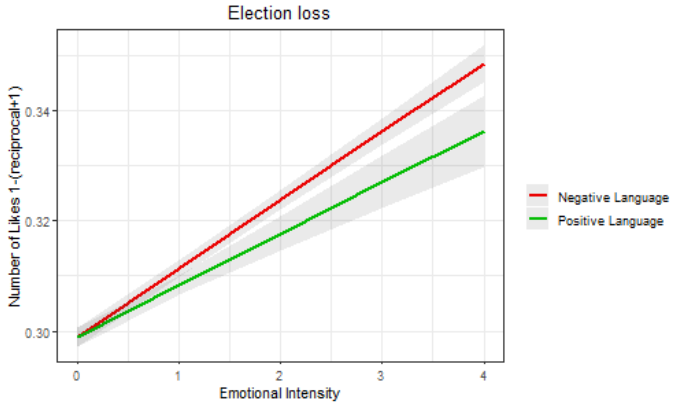
Model Fit

R^2	Marginal	Conditional
	0.066	0.38

Model equation: Likes 1-(reciprocal+1) ~ Positive * Negative + centered (Followers) + (1 | User)

Notes. *Model fit was calculated using the R package MuMIn (Barton, 2018) based on the paper of Nakagawa et al. (2017).*

Figure 3.1S. *Results of emotional language intensity (negative and positive) predicting number of likes (reciprocal+1 transformed) for the tweets mourning Hillary Clinton's election loss.*



Notes. *I found that an increase in negative language intensity was associated with an increase in the number of likes. Positive language intensity was also associated with an increase in likes.*

Trump election win

Table 3.2S. *Linear mixed model with four factors (positive language, negative language, number of followers, positive language x negative language) and number of likes 1-(reciprocal+1) as the dependent variable.*

Fixed Effects

	Estimate	*SE*	95% CI	*t*	*p*

Intercept	0.27	0.00098	0.27 – 0.27	276.47	.000	
Positive Language	0.0071	0.00074	0.0057 – 0.0085	9.61	.000	
Negative Language		-0.0015	0.00064	-0.0027 – -0.00030	-2.34	.019
Number of followers		0.097	0.00085	0.095 – 0.098	113.87	.000
Positive Language × Negative Language		-0.00071	0.00065	-0.0017 – 0.00049	-1.07	.280

Random Effects

	Variance	*SD*
Participant (Intercept)	0.038	0.20
Residual	0.044	0.21

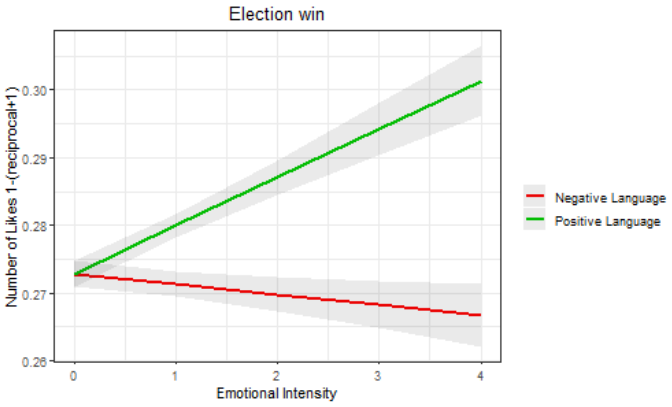
Model fit

R^2	Marginal	Conditional
	0.10	0.52

Model equation: Likes 1-(reciprocal+1) ~ Positive * Negative + centered (Followers) + (1 | User)

Notes. *Model fit was calculated using the R package MuMIn (Barton, 2018) based on the paper of Nakagawa et al. (2017).*

Figure 3.2S. *Results of emotional language intensity (negative and positive) predicting number of likes (reciprocal+1 transformed) expressed in the tweets celebrating Donald Trump's victory.*



Notes. *I found that an increase in negative language intensity was associated with a decrease in the number of retweets (only example where negative language was a negative predictor), while an increase in positive language was associated with an increase in retweets. This result highlights the difference between likes and retweets as negative language was consistently a positive predictor for an increase in number of retweets, while we saw a decrease in the number of likes in this particular event.*

Ferguson unrest

Table 3.3S. *Linear mixed model with four factors (positive language, negative language, number of followers, positive language x negative language) and number of likes 1-(reciprocal+1) as the dependent variable.*

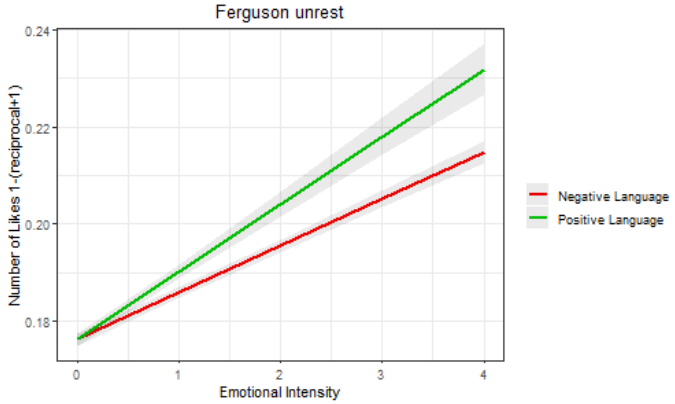
Fixed Effects					
	Estimate	*SE*	95% CI	*t*	*p*
Intercept	0.17	0.00066	0.17 – 0.18	277.83	.000
Positive Language	0.013	0.00071	0.012 – 0.014	19.77	.000
Negative Language	0.0096	0.00036	0.0087 – 0.010	26.76	.000
Number of followers	0.079	0.00050	0.078 – 0.080	156.33	.000
Positive Language × Negative Language	-0.0017	0.00050	-0.0029 – 0.00082	-3.45	.000
Random Effects					

	Variance	*SD*
Participant (Intercept)	0.013	0.12
Residual	0.051	0.22

Model equation: Likes 1-(reciprocal+1) ~ Positive * Negative + centered (Followers) + (1 | User)

Notes. *Model fit was calculated using the R package MuMIn (Barton, 2018) based on the paper of Nakagawa et al. (2017).*

Figure 3.3S. *Results of emotional language intensity (negative and positive) predicting number of likes (reciprocal+1 transformed) expressed in the Ferguson unrest tweets.*



Notes. *I found that an increase in negative language intensity was associated with an increase in the number of likes.*

Same-sex ruling

Table 3.4S. *Linear mixed model with four factors (positive language, negative language, number of followers, positive language x negative language) and number of likes 1-(reciprocal+1) as the dependent variable.*

Fixed Effects

	Estimate	*SE*	95% CI	*t*	*p*
Intercept	0.24	0.00064	0.23 – 0.23	381.09	.000
Positive Language	0.0077	0.00044	0.0052 – 0.0070	17.40	.000

Negative Language	0.0086	0.00063	0.0055 – 0.0085	13.64	.000
Number of followers	0.098	0.00045	0.094 – 0.099	218.21	.000
Positive Language × Negative Language	0.0024	0.00048	-0.0012 – 0.0034	5.02	.000

Random Effects

	Variance	*SD*
Participant (Intercept)	0.031	0.17
Residual	0.047	0.21

Model fit

Model equation: Likes 1-(reciprocal+1) ~ Positive * Negative + centered (Followers) + (1 | User)

Notes. *Model fit was calculated using the R package MuMIn (Barton, 2018) based on the paper of Nakagawa et al. (2017).*

Figure 3.4S. *Results of emotional language intensity (negative and positive) predicting number of likes (reciprocal+1 transformed) expressed in the same-sex marriage tweets.*



Notes. *I found that an increase in negative language intensity was associated with an increase in the number of likes. Positive language intensity was also associated with an increase in likes.*

Emotional Language Intensity Predicting Retweets in Subsets for Positive Situations

I repeated the main analysis of the manuscript that assessed the spread of content for both positive situations in each study (Study 1: Election win, Study 2: Same-sex marriage) with a subsample that only contained users that belonged to the political group that supported the event as estimated by our political affiliation estimation. This analysis addressed the concern that the effect of negative language on spreadibility of content was caused by an opposing political group expressing their negative language content.

Trump election win (Subset 22,642 tweets from conservatives)

Table 3.5S. *Linear mixed model with four factors (positive language, negative language, number of followers, positive language x negative language) and number of retweets 1-(reciprocal+1) as the dependent variable.*

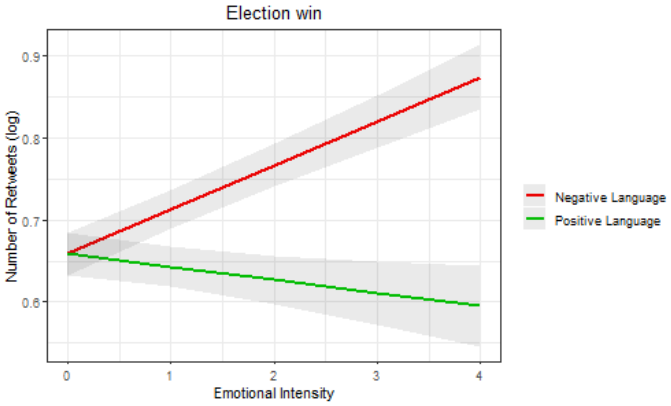
Fixed Effects					
	Estimate	*SE*	95% CI	*t*	*p*
Intercept	0.27	0.0042	0.27 – 0.28	66.40	.000
Positive Language	-0.0059	0.0026	-0.010 – -0.0013	-2.29	.021
Negative Language	0.031	0.0020	0.026 – 0.035	15.24	.000
Number of followers	0.071	0.0033	0.065 – 0.077	21.20	.000
Positive Language × Negative Language	0.0059	0.0020	0.0025 – 0.010	2.89	.003
Random Effects					
			Variance	*SD*	
Participant (Intercept)			0.40	0.63	
Residual			0.29	0.54	
Model fit					
R^2			Marginal	Conditional	

	0.071	0.40

Model equation: Retweets 1-(reciprocal+1)~ Positive * Negative + centered (Followers) + (1 | User)

Notes. *Model fit was calculated using the R package MuMIn (Barton, 2018) based on the paper of Nakagawa et al. (2017).*

Figure 3.5S. *Results of emotional language intensity (negative and positive) predicting number of retweets (1-(reciprocal+1) transformed) expressed in the tweets celebrating Donald Trump's victory (subsample of estimated conservatives).*



Notes. *I found that an increase in negative language intensity was associated with an increase in the number of retweets, while an increase in positive language was negatively correlated with the number of retweets.*

Same-sex ruling (Subset 7.661 tweets from liberals)

Table 3.6S. *Linear mixed model with four factors (positive language, negative language, number of followers, positive language x negative language) and number of retweets 1-(reciprocal+1) as the dependent variable.*

Fixed Effects

	Estimate	*SE*	95% CI	*t*	*p*
Intercept	0.38	0.011	0.35 – 0.40	33.77	.000
Positive Language	0.0026	0.0073	-0.0091 – 0.017	0.35	.723
Negative Language	0.060	0.0074	0.042 – 0.079	8.17	.000

Number of followers	0.13	0.0083	0.12 – 0.15	16.25	.000
Positive Language × Negative Language	0.0039	0.0056	-0.0087 – 0.014	0.69	.485

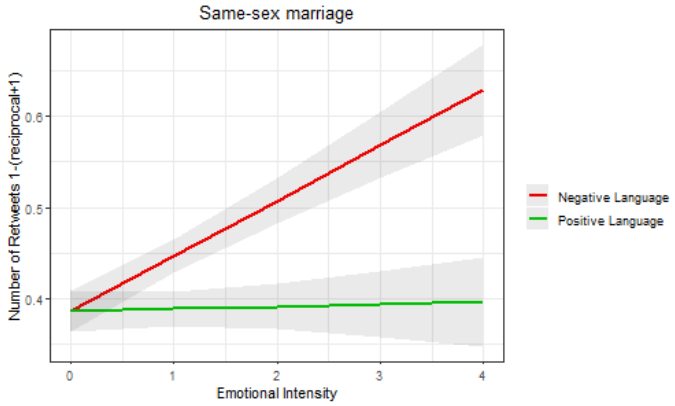
Random Effects		
	Variance	*SD*
Participant (Intercept)	0.017	0.13
Residual	0.052	0.22

Model fit		
R^2	Marginal	Conditional
	0.20	0.40

Model equation: Retweets 1-(reciprocal+1) ~ Positive * Negative + centered (Followers) + (1 | User)

Notes. *Model fit was calculated using the R package MuMIn (Barton, 2018) based on the paper of Nakagawa et al. (2017).*

Figure 3.6S. *Results of emotional language intensity (negative and positive) predicting number of retweets (1-(reciprocal+1) transformed) expressed in the same-sex marriage tweets (subsample of estimated liberals).*



Notes. *I found that an increase in negative language intensity is associated with an increase in the number of retweets, while an increase in positive language was not correlated with number of retweets.*

Emotional Language Intensity Predicting Retweets: Model Comparisons

I tested three different models that could substitute the analysis presented in the manuscript. I compared (a) a linear models with 1-(reciprocal)+1 transformed retweet counts against (b) a linear models with log+1 transformed retweet counts (model from the manuscript), (c) a Poisson regression model, and (d) a negative binomial model. Based on the direct model comparison we decided to use the linear models with reciprocal+1 transformed retweet counts as it presented the best model fit in predicting retweets by the positive and negative sentiment of the text.

Model Details:

a) Linear regression 1-(reciprocal+1): This model is the one we originally presented in the paper. Our reciprocal transformation, is denoted by $f(x) = 1-(x+1)^{-1}$. In simple words, reciprocal transformation involves dividing 1 by the retweet number. Since many of our between numbers were zeros, we first conducted a x+1 transformation. Finally, while reciprocal transformation reduces the skewness in the data it also reveres the numbers' magnitude. To account for this, we subtracted the number from 1. It is noteworthy that the distribution is still not strictly normally distributed due to the excess zeros. The model itself remained the same as in the manuscript meaning that we predicted the number of retweets using both the positive and the negative emotional language score for each tweet, as well as their interaction. The interaction term was introduced to the model in order to assess whether the effects of either content valence on spread depended on the level of the other content valence (thus testing for a possible effect of mixed emotional language). I again added a random intercept of user id to the model as some users wrote more than one tweet.

   Lmer (reciprocal(retweet+1) ~

   positive_score + negative_score + positive_score *negative_score + (user_id),

   family = "gaussian")

b) Linear regression log-transform: I log+1 transformed the number of retweets. The distribution was still severely skewed and had excess zeros, leading to a positive skew

of the distribution. The model's variables are identical to the ones from the previous model (a).

Lmer (log(retweet+1) ~

positive_score + negative_score + positive_score *negative_score + (user_id),

family = "gaussian")

c) Poisson regression model: In an attempt to deal with the excess zeros we also considered a Poisson count model. The model's variables were the same as previously specified in A.

Lmer (retweet ~

positive_score + negative_score + positive_score *negative_score + (user_id),

family = "Poisson")

d) Negative binomial regression: Similar to the Poisson regression, we checked whether a negative binomial count model is better able to accommodate the large amount of zeros leading to a positive skew. The model's variables are identical to the ones from the previous model (A.).

Lmer.nb (retweet ~

positive_score + negative_score + positive_score *negative_score + (user_id),

family = "negative binomial")

Prediction of models: All models for all situations in the studies show similar results, finding that negativity spreads further than positivity. The only exception is the negative binomial model for the election win dataset, which did not show a significant association between positive and negative emotion scores of tweets and the number of retweets.

Trump election win

Table 3.7S. Linear regression 1-(reciprocal+1) model: *Linear mixed model with four factors (positive language, negative language, number of followers, positive language x negative language) and number of retweets 1-(reciprocal+1) as the dependent variable.*

Fixed Effects					
	Estimate	*SE*	95% CI	*t*	*p*
Intercept	0.13	0.00077	0.13 – 0.13	167.80	.000
Positive Language	0.0003	0.00062	- 0.000811– 0.0016	0.63	.52
Negative Language	0.0018	0.00054	0.00084 – 0.0028	3.44	.000
Number of followers	0.0808	0.00064	0.079 – 0.082	124.61	.000
Positive Language × Negative Language	-0.00068	0.00055	- 0.0019 – 0.00049	-1.22	.219
Random Effects					
			Variance	SD	
Participant (Intercept)			0.017	0.13	
Residual			0.034	0.18	
Model fit					
R^2			Marginal	Conditional	
			0.12	0.41	

Model equation: Retweets 1-(reciprocal+1) ~ Positive * Negative + centered (Followers) + (1 | User)

Notes. *Model fit was calculated using the R package MuMIn (Barton, 2018) based on the paper of Nakagawa et al. (2017).*

Table 3.8S. Linear regression (log+1) model: *Linear mixed model with four factors (positive language, negative language, number of followers, positive language x negative language) and number of retweets (log+1) as the dependent variable.*

Fixed Effects					
	Estimate	*SE*	95% CI	*t*	*p*
Intercept	0.21	0.0015	0.35 – 0.40	137.62	.000
Positive Language	0.0024	0.0012	-0.0052 – 0.0046	2.019	.043
Negative Language	0.0046	0.0010	0.0023 – 0.0065	4.33	.000
Number of followers	0.16	0.0013	0.16 – 0.17	124.28	.000
Positive Language × Negative Language	-0.0018	0.0010	-0.0041– 0.00038	-1.72	.085

Random Effects		
	Variance	*SD*
Participant (Intercept)	0.090	0.30
Residual	0.12	0.35

Model fit		
R^2	Marginal	Conditional
	0.11	0.48

Model equation: Retweets (log+1) ~ Positive * Negative + centered (Followers) + (1 | User)

Notes. Model fit was calculated using the R package MuMIn (Barton, 2018) based on the paper of Nakagawa et al. (2017).

Table 3.9S. Poisson regression model: *Count mixed model with four factors (positive language, negative language, number of followers, positive language x negative language) and number of retweets as the dependent variable.*

Fixed Effects					
	Estimate	*SE*	95% CI	*t*	*p*

Intercept	-2.23	0.010	-2.31– -2.16	-213.30	.000
Positive Language	0.15	0.0023	-0.0010 – 0.0025	64.31	.000
Negative Language	0.0084	0.0022	-0.0045 – 0.019	3.74	.000
Number of followers	0.92	0.0082	0.91 – 0.95	112.22	.000
Positive Language x Negative Language	-0.11	0.0024	-0.016 – 0.0078	-47.06	.000
Random Effects					
			Variance	*SD*	
Participant (Intercept)			2.19	1.48	
Model fit					
R^2			Marginal	Conditional	
			0.09	0.33	

Model equation: Retweets ~ Positive * Negative + centered (Followers) + (1 | User), family = Poisson

Notes. *Model fit was calculated using the R package MuMIn (Barton, 2018) based on the paper of Nakagawa et al. (2017).*

Table 3.10S. Negative binomial regression model: *Count mixed model with four factors (positive language, negative language, number of followers, positive language x negative language) and number of retweets as the dependent variable.*

Fixed Effects					
	Estimate	*SE*	95% CI	*t*	*p*
Intercept	-2.10	0.012	-2.16 – -2.31	-175.50	.000
Positive Language	0.013	0.0069	-0.010 – 0.025	1.86	.616
Negative Language	0.0083	0.0060	-0.045 – 0.019	1.37	.168

Number of followers	0.95	0.0083	0.91 – 0.99	114.40	.000
Positive Language × Negative Language	-0.0056	0.0062	-0.016 – 0.0078	-0.905	.365
Random Effects					
			Variance	*SD*	
Participant (Intercept)			1.88	1.37	
Model fit					
R^2			Marginal	Conditional	
			0.03	0.09	

Model equation: Retweets ~ Positive * Negative + centered (Followers) + (1 | User), family = Poisson

Notes. *Model fit was calculated using the R package MuMIn (Barton, 2018) based on the paper of Nakagawa et al. (2017).*

Table 3.11S. Model Comparison: *Comparing log-likelihood scores of the above described models.*

	Marginal R^2	AIC	BIC	Log-Likelihood	df	chi-squared	*p*
1 - (Reciprocal + 1) Model	.12	-76956	-76882	38452	7		
Log+1 Model	.11	303874	303947	-151959	7	380823	.000
Poisson Model	.09	979581	979645	-489785	6	675651	.000
Negative Binomial Model	.03	411925	411999	-205956	7	567658	.000

Notes. *Model fit was calculated using the R package lmtest (Hothorn et al., 2015)*

Hillary election loss

Table 3.12S. Linear regression 1-(reciprocal+1) model: *Linear mixed model with four factors (positive language, negative language, number of followers, positive language x negative language) and number of retweets 1-(reciprocal+1) as the dependent variable.*

Fixed Effects

	Estimate	*SE*	95% CI	*t*	*p*
Intercept	0.16	0.00077	0.16 – 0.16	208.34	.000
Positive Language	0.0010	0.00080	-0.00072 – 0.0025	1.25	.209
Negative Language	0.0060	0.00045	0.0049 – 0.0069	13.45	.000
Number of followers	0.072	0.00059	0.070 – 0.073	121.88	.000
Positive Language × Negative Language	0.00065	0.00054	-0.00031– 0.0019	1.21	.219

Random Effects		
	Variance	*SD*
Participant (Intercept)	0.019	0.14
Residual	0.046	0.21

Model fit		
R^2	Marginal	Conditional
	0.07	0.35

Model equation: Retweets 1-(reciprocal+1) ~ Positive * Negative + centered (Followers) + (1 | User)

Notes. *Model fit was calculated using the R package MuMIn (Barton, 2018) based on the paper of Nakagawa et al. (2017).*

Table 3.13S. Linear regression (log+1) model: *Linear mixed model with four factors (positive language, negative language, number of followers, positive language x negative language) and number of retweets (log+1) as the dependent variable.*

Fixed Effects					
	Estimate	*SE*	95% CI	*t*	*p*
Intercept	0.092	0.00056	0.091 – 0.093	163.32	.000
Positive Language	0.00057	0.00056	-0.00046 – 0.0017	1.02	.309

Negative Language	0.0037	0.00031	0.0031 – 0.0043	12.05	.000
Number of followers	0.054	0.00045	0.053 – 0.055	120.23	.000
Positive Language × Negative Language	0.00015	0.00037	-0.00053 – 0.00094	0.40	.682

Random Effects		
	Variance	*SD*
Participant (Intercept)	0.014	0.12
Residual	0.02	0.14

Model fit		
R^2	Marginal	Conditional
	0.08	0.45

Model equation: Retweets (log+1) ~ Positive * Negative + centered (Followers) + (1 | User)

Notes. *Model fit was calculated using the R package MuMIn (Barton, 2018) based on the paper of Nakagawa et al. (2017).*

Table 3.14S. Poisson regression model: *Count mixed model with four factors (positive language, negative language, number of followers, positive language x negative language) and number of retweets as the dependent variable.*

Fixed Effects					
	Estimate	*SE*	95% CI	*t*	*p*
Intercept	-1.60	0.0073	-1.69 – -1.55	-219.94	.000
Positive Language	-0.11	0.0028	-0.12 – -0.10	-40.36	.000
Negative Language	-0.015	0.0058	-0.020 – -0.0087	-9.74	.000
Number of followers	0.52	0.0018	0.50 – 0.53	89.43	.000

Positive Language × Negative Language	0.020	0.0024	0.012 – 0.027	-10.84	.000
Random Effects					
			Variance	SD	
Participant (Intercept)			2.23	1.49	
Model fit					
R^2			Marginal	Conditional	
			0.05	0.44	

Model equation: Retweets ~ Positive * Negative + centered (Followers) + (1 | User), family = Poisson

Notes. Model fit was calculated using the R package MuMIn (Barton, 2018) based on the paper of Nakagawa et al. (2017).

Table 3.15S. Negative binomial regression model: *Count mixed model with four factors (positive language, negative language, number of followers, positive language x negative language) and number of retweets as the dependent variable.*

Fixed Effects					
	Estimate	*SE*	95% CI	*t*	*p*
Intercept	-1.61	0.0089	-1.64 --1.59	-181.51	.000
Positive Language	0.0050	0.0077	-0.017 - 0.023	0.65	.514
Negative Language	0.054	0.0042	0.051 - 0.058	12.68	.000
Number of followers	0.67	0.0060	0.64 - 0.66	113.16	.000
Positive Language × Negative Language	0.0035	0.0050	-0.0050 - 0.017	0.69	.490
Random Effects					
			Variance	*SD*	
Participant (Intercept)			1.76	1.32	

Model fit		
R^2	Marginal	Conditional
	.03	.16

Model equation: Retweets ~ Positive * Negative + centered (Followers) + (1 | User), family = Poisson

Notes. *Model fit was calculated using the R package MuMIn (Barton, 2018) based on the paper of Nakagawa et al. (2017).*

Table 3.16S. Model Comparison: *Comparing log-likelihood scores of the above described models.*

	Marginal R^2	AIC	BIC	Log-Likelihood	df	chi-squared	*p*
1 - (Reciprocal + 1) Model	.07	2706	2781	-1379	7		
Log+1 Model	.08	-241976	303947	120960	7	244679	.000
Poisson Model	.05	1467897	1467961	-733942	6	1709805	.000
Negative Binomial Model	0.3	651565	651640	-325775	7	816334	.000

Notes. *Model fit was calculated using the R package lmtest (Hothorn et al., 2015)*

Same-sex ruling

Table 3.17S. Linear regression 1-(reciprocal+1) model: *Linear mixed model with four factors (positive language, negative language, number of followers, positive language x negative language) and number of retweets 1-(reciprocal+1) as the dependent variable.*

Fixed Effects					
	Estimate	*SE*	95% CI	*t*	*p*
Intercept	0.12	0.00050	0.12 – 0.12	241.28	.000
Positive Language	0.0014	0.00035	0.00075 – 0.0020	4.02	.000
Negative Language	0.0063	0.00051	0.0055 – 0.0075	12.43	.000
Number of followers	0.097	0.00034	0.096 - 0.097	283.07	.000

Positive Language × Negative Language	0.00057	0.00039	-0.00024 - 0.0012	1.45	.145
Random Effects					
			Variance	*SD*	
Participant (Intercept)			0.013	0.11	
Residual			0.034	0.18	
Model fit					
R^2			Marginal	Conditional	
			.16	.40	

Model equation: Retweets 1-(reciprocal+1) ~ Positive * Negative + centered (Followers) + (1 | User)

Notes. *Model fit was calculated using the R package MuMIn (Barton, 2018) based on the paper of Nakagawa et al. (2017).*

Table 3.18S. Linear regression (log+1) model: *Linear mixed model with four factors (positive language, negative language, number of followers, positive language x negative language) and number of retweets (log+1) as the dependent variable.*

Fixed Effects					
	Estimate	*SE*	95% CI	*t*	*p*
Intercept	0.21	0.0010	0.21 – 0.21	201.85	.000
Positive Language	0.0035	0.00072	0.0023 - 0.0050	4.93	.309
Negative Language	0.0097	0.0010	0.0078 - 0.011	9.31	.000
Number of followers	0.22	0.00072	0.22 - 0.22	307.05	.000
Positive Language × Negative Language	0.0010	0.00079	-0.00046 - 0.0022	1.36	.172
Random Effects					
			Variance	*SD*	

Participant (Intercept)	0.0085	0.092
Residual	0.014	0.12

Model fit		
R^2	Marginal	Conditional
	.19	.49

Model equation: Retweets (log+1) ~ Positive * Negative + centered (Followers) + (1 | User)

Notes. *Model fit was calculated using the R package MuMIn (Barton, 2018) based on the paper of Nakagawa et al. (2017).*

Table 3.19S. Poisson regression model: *Count mixed model with four factors (positive language, negative language, number of followers, positive language x negative language) and number of retweets as the dependent variable.*

Fixed Effects					
	Estimate	SE	95% CI	*t*	*p*
Intercept	-2.17	0.0062	-2.17 – -2.16	-349.68	.000
Positive Language	-0.065	0.0016	0.070 - -0.060	-38.97	.000
Negative Language	-0.15	0.0030	-0.15 - -0.14	-50.50	.000
Number of followers	1.09	0.0041	1.08 - 1.10	262.14	.000
Positive Language × Negative Language	0.17	0.0023	0.17 - 0.17	73.70	.000

Random Effects		
	Variance	*SD*
Participant (Intercept)	1.95	1.40

Model fit		
R^2	Marginal	Conditional

	.14	.39

Model equation: Retweets ~ Positive * Negative + centered (Followers) + (1 | User), family = Poisson

Notes. *Model fit was calculated using the R package MuMIn (Barton, 2018) based on the paper of Nakagawa et al. (2017).*

Table 3.20S. Negative binomial regression model: *Count mixed model with four factors (positive language, negative language, number of followers, positive language x negative language) and number of retweets as the dependent variable.*

Fixed Effects					
	Estimate	*SE*	95% CI	*t*	*p*
Intercept	-2.16	0.0077	-2.17 - -2.16	-280.39	.000
Positive Language	0.0050	0.0042	-0.0036 - 0.012	1.19	.232
Negative Language	0.054	0.0058	0.048 - 0.058	9.24	.000
Number of followers	1.081	0.0041	1.054 - 1.076	259.49	.000
Positive Language × Negative Language	0.012	0.0044	0.0091 - 0.017	2.75	.006
Random Effects					
			Variance	*SD*	
Participant (Intercept)			1.60	1.26	
Model fit					
R^2			Marginal	Conditional	
			.027	.06	

Model equation: Retweets ~ Positive * Negative + centered (Followers) + (1 | User), family = Poisson

Notes. *Model fit was calculated using the R package MuMIn (Barton, 2018) based on the paper of Nakagawa et al. (2017).*

Table 3.21S. Model Comparison: *Comparing log-likelihood scores of the above described models.*

	Marginal R^2	AIC	BIC	Log-Likelihood	df	chi-squared	*p*
1 - (Reciprocal + 1) Model	.16	-139750	-139671	69882	7		
Log+1 Model	.19	-572045	-571966	286029	7	432291	.000
Poisson Model	.15	1430343	1430411	-715166	6	2002316	.000
Negative Binomial Model	.027	805288	805367	-402637	7	625057	.000

Notes. *Model fit was calculated using the R package lmtest (Hothorn et al., 2015)*

Ferguson unrest

Table 3.22S. Linear regression 1-(reciprocal+1) model: *Linear mixed model with four factors (positive language, negative language, number of followers, positive language x negative language) and number of retweets 1-(reciprocal+1) as the dependent variable.*

Fixed Effects					
	Estimate	*SE*	95% CI	*t*	*p*
Intercept	0.18	0.00064	0.17 – 0.18	279.78	.000
Positive Language	-0.0011	0.00072	-0.0026 - 0.00019	-1.53	.124
Negative Language	0.0079	0.00036	0.0071 - 0.0087	21.56	.000
Number of followers	0.099	0.00051	0.097 - 0.10	192.47	.000
Positive Language × Negative Language	-0.00054	0.00051	-0.0016 - 0.00026	-1.05	.291

Random Effects		
	Variance	*SD*
Participant (Intercept)	0.013	0.12
Residual	0.054	0.23

Model fit		
R^2	Marginal	Conditional
	.13	.29

Model equation: Retweets 1-(reciprocal+1) ~ Positive * Negative + centered (Followers) + (1 | User)

Notes. *Model fit was calculated using the R package MuMIn (Barton, 2018) based on the paper of Nakagawa et al. (2017).*

Table 3.23S. Linear regression (log+1) model: *Linear mixed model with four factors (positive language, negative language, number of followers, positive language x negative language) and number of retweets (log+1) as the dependent variable.*

Fixed Effects					
	Estimate	*SE*	95% CI	*t*	*p*
Intercept	0.30	0.0013	0.30 – 0.30	221.32	.000
Positive Language	-0.0046	0.0015	-0.0083 - -0.0021	-3.09	.002
Negative Language	0.016	0.00077	0.015 - 0.018	21.89	.000
Number of followers	0.21	0.0011	0.21 - 0.21	191.53	.000
Positive Language x Negative Language	-0.000032	0.0010	-0.0024 - 0.0019	-0.03	.976

Random Effects		
	Variance	*SD*
Participant (Intercept)	0.07	0.26
Residual	0.23	0.48

Model fit		
R^2	Marginal	Conditional
	.13	.33

Model equation: Retweets (log+1) ~ Positive * Negative + centered (Followers) + (1 | User)

Notes. *Model fit was calculated using the R package MuMIn (Barton, 2018) based on the paper of Nakagawa et al. (2017).*

Table 3.24S. Poisson regression model: *Count mixed model with four factors (positive language, negative language, number of followers, positive language x negative language) and number of retweets as the dependent variable.*

Fixed Effects					
	Estimate	*SE*	95% CI	*t*	*p*
Intercept	-1.39	0.0055	-1.39 – -1.38	- 253.52	.000
Positive Language	-0.073	0.0025	-0.083 - -0.061	-28.68	.000
Negative Language	0.063	0.0011	0.060 - 0.067	52.95	.000
Number of followers	0.73	0.0040	0.72 - 0.74	183.09	.000
Positive Language × Negative Language	0.036	0.0017	0.031 - 0.040	20.69	.000
Random Effects					
			Variance	*SD*	
Participant (Intercept)			1.39	1.18	
Model fit					
R^2			Marginal	Conditional	
			.09	.33	

Model equation: Retweets ~ Positive * Negative + centered (Followers) + (1 | User), family = Poisson

Notes. *Model fit was calculated using the R package MuMIn (Barton, 2018) based on the paper of Nakagawa et al. (2017).*

Table 3.25S. Negative binomial regression model: *Count mixed model with four factors (positive language, negative language, number of followers, positive language x negative language) and number of retweets as the dependent variable.*

a.) Fixed Effects

	Estimate	*SE*	95% CI	*t*	*p*
Intercept	-1.27	0.0064	-1.28 - -1.26	-197.14	.000
Positive Language	-0.014	0.0059	-0.020 - -0.0053	-2.42	.015
Negative Language	0.065	0.0029	0.055 - 0.070	21.84	.000
Number of followers	0.85	0.0048	0.82 - 0.84	177.44	.000
Positive Language × Negative Language	-0.0032	0.0042	-0.011 - -0.00075	-0.76	.445

Random Effects

	Variance	*SD*
Participant (Intercept)	0.99	0.99

Model fit

R^2	Marginal	Conditional
	.04	.09

Model equation: Retweets ~ Positive * Negative + centered (Followers) + (1 | User), family = Poisson

Notes. *Model fit was calculated using the R package MuMIn (Barton, 2018) based on the paper of Nakagawa et al. (2017).*

Table 3.26S. Model Comparison: *Comparing log-likelihood scores of the above described models.*

	Marginal R^2	AIC	BIC	Log-Likelihood	df	chi-squared	*p*
1-(Reciprocal + 1) Model	.13	40115	40193	-20084	7		

Log+1 Model	.13	862078	862157	-431062	7	821956	.000
Poisson Model	.09	2567482	2567550	-1283735	6	1705346	.000
Negative Binomial Model	.04	1180476	1180554	-590231	7	1387009	.000

Notes. *Model fit was calculated using the R package lmtest (Hothorn et al., 2015)*

Emotional Language Intensity Predicting Retweets: General Additive Models

To test if linear parametric models are the appropriate choice I ran general additive mixed model via mgcv package in R (Wood, 2015). As in the linear mixed model, I predicted the 1-(reciprocal+1) transformed retweet count using positive and negative SentiStrength scores. The number of knots for both positive and negative scores were 4. The results supported the initial claim that negativity spreads further than positivity in all situations.

Trump election win

Table 3.27S. GAM 1-(reciprocal+1) model: *General additive model with two smooth factors (positive language, negative language) using 4 knots and one parametric factor (number of followers) predicting number of retweets 1-(reciprocal+1) as the dependent variable.*

Non-Parametric Terms				
	Estimate	*SE*	*t*	*p*
Intercept	-0.098	0.0018	-53.38	.000
Number of followers	0.039	0.00031	124.61	.000
Smooth-Terms				
	edf	Ref.df	*F*	*p*
Positive Language (4 knots)	1.00	1.00	1.25	.884
Negative Language (4 knots)	2.443	2.443	6.86	.000
Random Effects				
		Variance	*SD*	

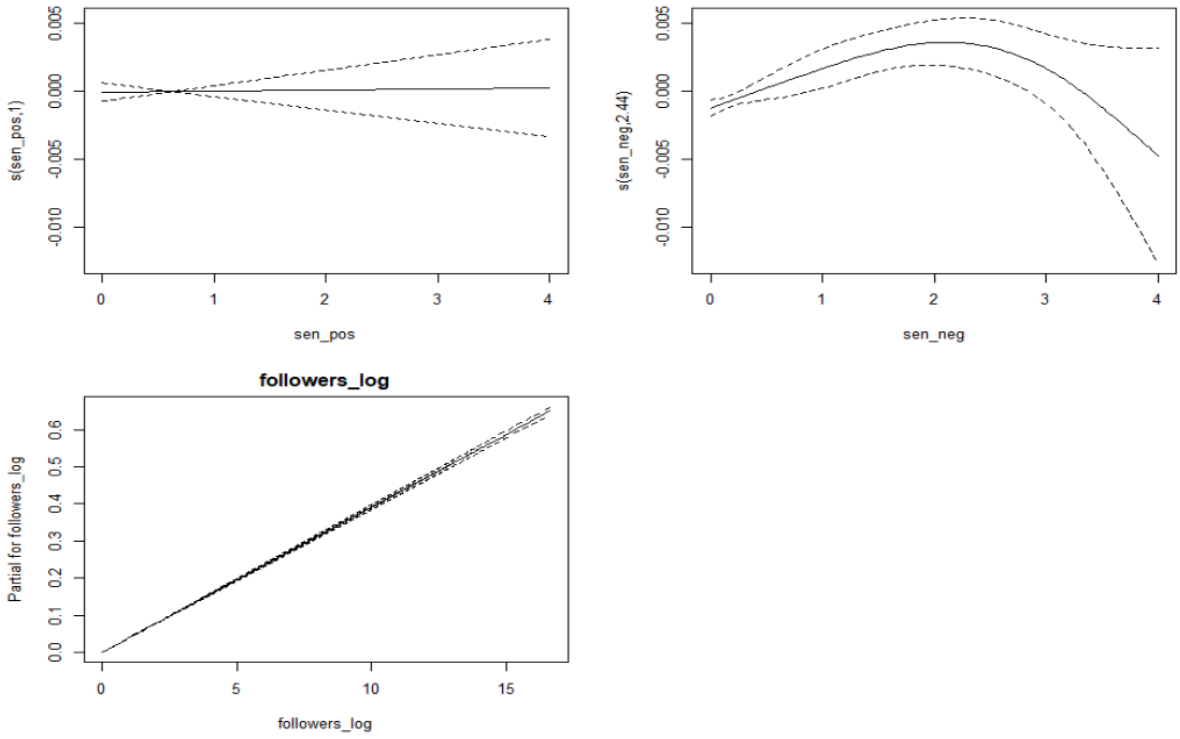
Participant (Intercept)	0.017	0.13
Residual	0.034	0.18

Model fit	
R^2	Adjusted
	.14

Model equation: Retweets 1-(reciprocal+1) ~ (Positive, knots = 4) + (Negative, knots = 4) + centered (Followers) + (1 | User)

Notes. *General additive mixed model was computed via mgcv package in R (Wood, 2015).*

Figure 3.7S. *Results of emotional language intensity (negative and positive) predicting number of retweets (1-(reciprocal+1) transformed) expressed in the election win dataset.*



Notes. *I found that an increase in negative language intensity is associated with an increase in the number of retweets, while an increase in positive language was not correlated with number of retweets. It is noteworthy, that for very intense negative scores the prediction is imprecise indicated by the large confidence intervals.*

Hillary election loss

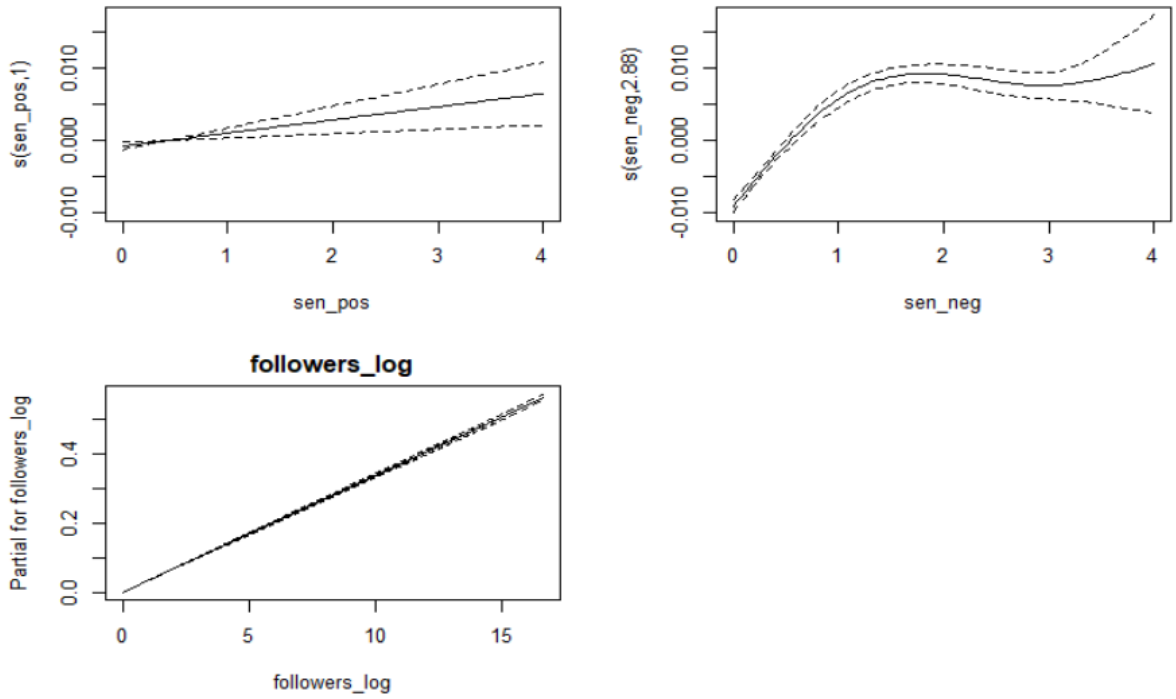
Table 3.28S. GAM 1-(reciprocal+1) model: *General additive model with two smooth factors (positive language, negative language) using 4 knots and one parametric factor (number of followers) predicting number of retweets 1-(reciprocal+1) as the dependent variable.*

Non-Parametric Terms				
	Estimate	*SE*	*t*	*p*
Intercept	-0.019	0.0016	-11.96	.000
Number of followers	0.033	0.00027	121.84	.000
Smooth-Terms				
	edf	Ref.df	*F*	*p*
Positive Language (4 knots)	1.00	1.00	8.92	.002
Negative Language (4 knots)	2.877	2.877	137.19	.000
Random Effects				
		Variance	*SD*	
Participant (Intercept)		0.046	0.21	
Residual		0.019	0.13	
Model fit				
R^2			Adjusted	
			.09	

Model equation: Retweets 1-(reciprocal+1) ~ (Positive, knots = 4) + (Negative, knots = 4) + centered (Followers) + (1 | User)

Notes. *General additive mixed model was computed via mgcv package in R (Wood, 2015).*

Figure 3.8S. *Results of emotional language intensity (negative and positive) predicting number of retweets (1-(reciprocal+1) transformed) expressed in the election win dataset.*



Notes. *I found that an increase in negative language intensity is associated with an increase in the number of retweets. An increase in positive language was also correlated with number of retweets.*

Same-sex marriage

Table 3.29S. GAM 1-(reciprocal+1) model: *General additive model with two smooth factors (positive language, negative language) using 4 knots and one parametric factor (number of followers) predicting number of retweets 1-(reciprocal+1) as the dependent variable.*

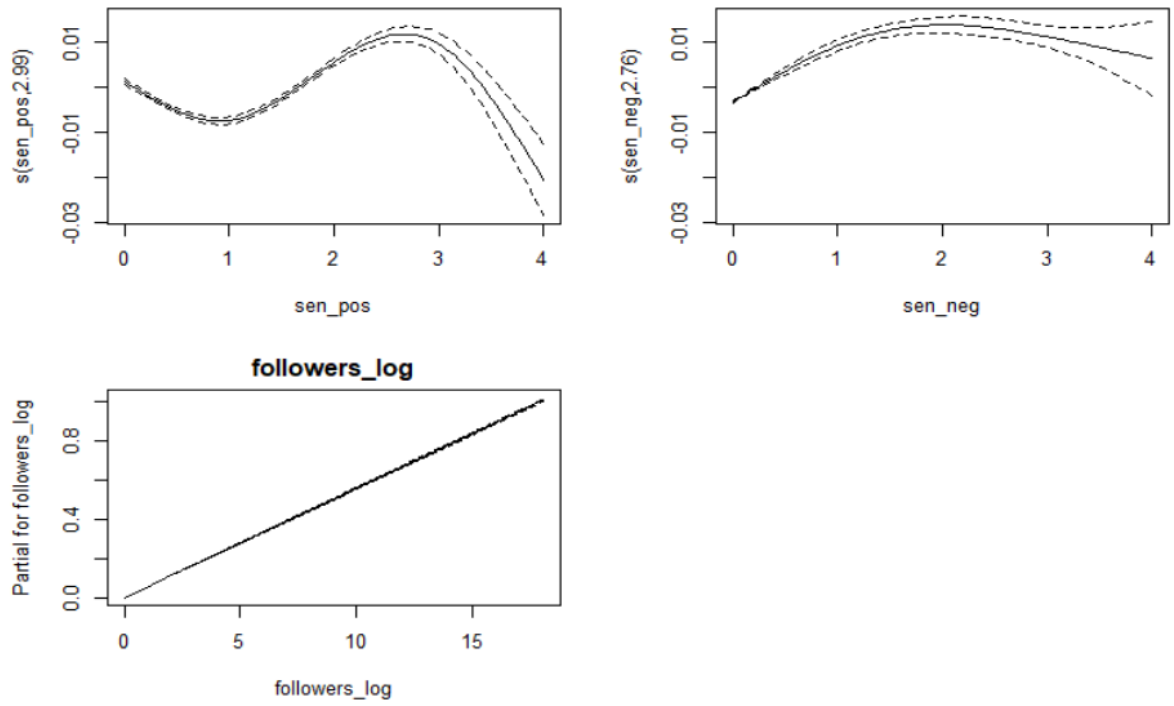
Non-Parametric Terms				
	Estimate	*SE*	*t*	*p*
Intercept	-0.22	0.0012	-181.0	.000
Number of followers	0.055	0.00019	283.2	.000
Smooth-Terms				
	edf	Ref.df	*F*	*p*
Positive Language (4 knots)	2.99	2.99	134.0	.002

Negative Language (4 knots)	2.75	2.75	154.7	.000
Random Effects				
		Variance	*SD*	
Participant (Intercept)		0.013	0.11	
Residual		0.034	0.18	
Model fit				
R^2			Adjusted	
			.18	

Model equation: Retweets 1-(reciprocal+1) ~ (Positive, knots = 4) + (Negative, knots = 4) + centered (Followers) + (1 | User)

Notes. *General additive mixed model was computed via mgcv package in R (Wood, 2015).*

Figure 3.9S. *Results of emotional language intensity (negative and positive) predicting number of retweets (1-(reciprocal+1) transformed) expressed in the election win dataset.*



Notes. *I found that an increase in negative language intensity is associated with an increase in the number of retweets. An increase in positive language was also correlated with number of retweets. However, it is important to note that the positive language score is not strictly linear suggesting that for very intense positive emotions there was a negative association with number of retweets.*

Ferguson unrest

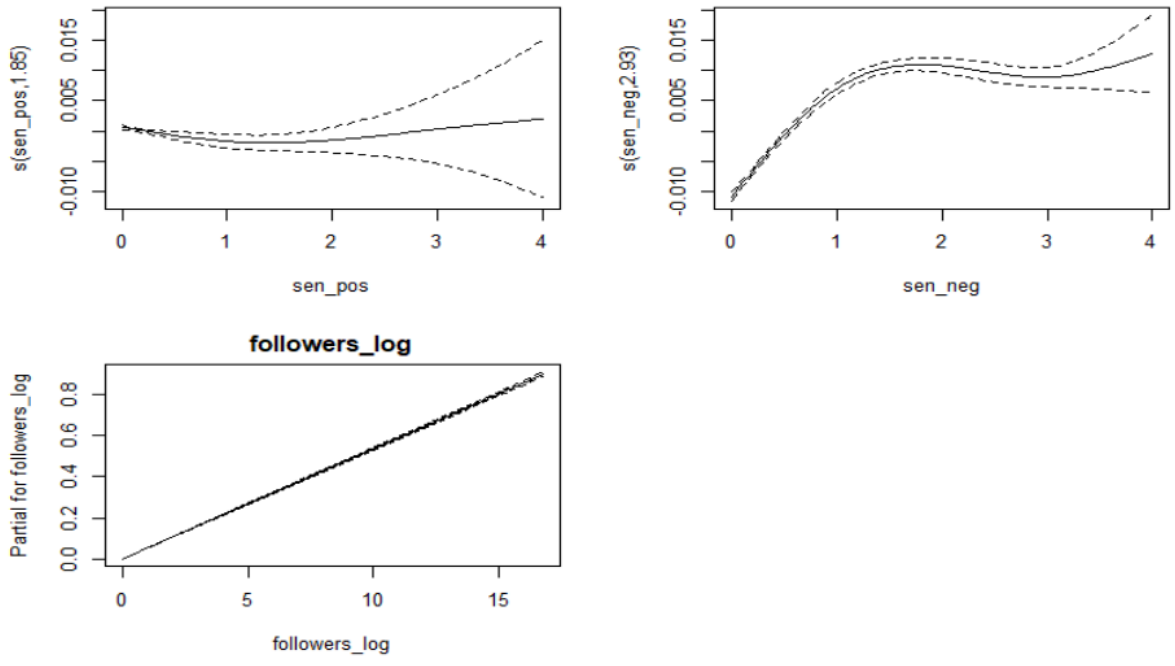
Table 3.30S. GAM 1-(reciprocal+1) model: *General additive model with two smooth factors (positive language, negative language) using 4 knots and one parametric factor (number of followers) predicting number of retweets 1-(reciprocal+1) as the dependent variable.*

a.) Non-Parametric Terms				
	Estimate	*SE*	*t*	*p*
Intercept	-0.14	0.0017	-86.33	.000
Number of followers	0.053	0.00027	192.41	.000
a.) Smooth-Terms				
	edf	Ref.df	*F*	*p*
Positive Language (4 knots)	1.85	1.85	3.66	.016
Negative Language (4 knots)	2.93	2.93	277.46	.000
Random Effects				
		Variance	*SD*	
Participant (Intercept)		0.013	0.11	
Residual		0.054	0.23	
Model fit				
R^2			Adjusted	
			.16	

Model equation: Retweets 1-(reciprocal+1) ~ (Positive, knots = 4) + (Negative, knots = 4) + centered (Followers) + (1 | User)

Notes. *General additive mixed model was computed via mgcv package in R (Wood, 2015).*

Figure 3.10S. *Results of emotional language intensity (negative and positive) predicting number of retweets (1-(reciprocal+1) transformed) expressed in the election win dataset.*



Notes. *I found that an increase in negative language intensity is associated with an increase in the number of retweets. An increase in positive language was not correlated with number of retweets. It is noteworthy, that for very intense negative scores the prediction is imprecise indicated by the large confidence intervals.*

Emotional Language Intensity Predicting Retweets: Sentiment Tool Comparison

To test whether SentiStrenght is the best choice of sentiment analysis tool for the political situations we chose, we compared SentiStrength against VADER (Gilbert & Hutto, 2014). I first correlated both scores.

a. Election Win (Positive Emotion r = 52, Negative Emotion r = .55),
b. Election Loss (Positive Emotion r = .55, Negative Emotion r =.62),
c. Same Sex Ruling (Positive Emotion r = .60, Negative Emotion r = .60),
d. Ferguson (Positive Emotion r = .53, Negative Emotion r = .63)

After that we ran the identical analysis as described above using the following formula:

Lmer (reciprocal(retweet+1) ~

positive_score + negative_vader + positive_vader *negative_vader + (user_id),

family = "gaussian")

Finally we compared the model fits using the log-likelihood test method from the lmtest package (Bates, Sarkar, Bates, & Matrix, 2007).

Trump election win

Table 3.31S. Linear regression 1-(reciprocal+1) model: *Linear mixed model with four factors (positive language, negative language, number of followers, positive language x negative language) and number of retweets 1-(reciprocal+1) as the dependent variable.*

	Fixed Effects				
	Estimate	*SE*	95% CI	*t*	*p*
Intercept	0.16	0.00080	0.13 - 0.13	165.31	.000

Positive Language Vader	-0.0074	0.0028	-0.013 - -0.0011	-2.56	.010
Negative Language Vader	0.0023	0.0041	-0.0055 - 0.010	0.57	.565
Number of followers	0.080	0.00064	0.079 - 0.081	124.49	.000
Positive Language V. x Negative Language V.	0.038	0.033	-0.0091 - 0.099	1.15	.249

Random Effects		
	Variance	*SD*
Participant (Intercept)	0.019	0.13
Residual	0.046	0.21

Model fit		
R^2	Marginal	Conditional
	.11	.41

Model equation: Retweets 1-(reciprocal+1) ~ Positive V. * Negative V. + centered (Followers) + (1 | User)

Notes. *Model fit was calculated using the R package MuMIn (Barton, 2018) based on the paper of Nakagawa et al. (2017).*

Table 3.32S. Model Comparison: *Comparing log-likelihood scores of SentiStrength vs. VADER models.*

	Marginal R^2	AIC	BIC	Log-Likelihood	df	chi-squared	*p*
SentiStrength	.12	-76956	-76882	38452	7		
VADER	.11	-76954	-76880	38459	7	13.097	.000

Notes. *Model fit was calculated using the R package lmtest (Hothorn et al., 2015)*

Hillary election loss

Table 3.33S. Linear regression 1-(reciprocal+1) model: *Linear mixed model with four factors (positive language, negative language, number of followers, positive language x negative language) and number of retweets 1-(reciprocal+1) as the dependent variable.*

Fixed Effects					
	Estimate	*SE*	95% CI	*t*	*p*
Intercept	0.16	0.00080	0.16 - 0.16	204.15	.000
Positive Language Vader	-0.0053	0.0028	-0.012 - 0.0017	-1.40	.160
Negative Language Vader	0.016	0.0041	0.0091 - 0.023	4.94	.000
Number of followers	0.072	0.00064	0.070 - 0.073	121.68	.000
Positive Language V. × Negative Language V.	0.24	0.028	0.18 - 0.29	8.68	.000

Random Effects		
	Variance	SD
Participant (Intercept)	0.019	0.14
Residual	0.047	0.22

Model fit		
R^2	Marginal	Conditional
	.07	.34

Model equation: Retweets 1-(reciprocal+1) ~ Positive V. * Negative V. + centered (Followers) + (1 | User)

Notes. *Model fit was calculated using the R package MuMIn (Barton, 2018) based on the paper of Nakagawa et al. (2017).*

Table 3.34S. Model Comparison: *Comparing log-likelihood scores of SentiStrength vs. VADER models.*

	Marginal R^2	AIC	BIC	Log-Likelihood	df	chi-squared	*p*
SentiStrength	.07	2706	2781.3	-1379.1	7		
VADER	.07	2706	2884.7	-1423.2	7	88.12	.000

Notes. *Model fit was calculated using the R package lmtest (Hothorn et al., 2015)*

Summary and Evaluation of Sentiment Tools for these Datasets.

The results for this analysis using VADER are identical in direction for the negative situations. The results for the election win were in the same direction to those reported in the manuscript but were not significant. The model fit suggested that the SentiStrength model were better in terms of their AIC and loglikelihood. As for the same sex marriage data, results did suggest that decreased negativity predicted retweets, but that positivity was also a significant negative predictor. Here again, our SentiStrength model was a better fit. One reason for VADER´s lower fit could be the way that positive and negative scores are computed. Vader has a positive, negative and a neutral score that are dependent on each other. These scores have to add up to 1, meaning that if a text contains many positive and negative words the score has to be divided between the three categories. I argue that especially in contexts that can feature both positive and negative emotions at the same time, rating positivity and negativity separately, as done by SentiStrength is crucial. The above analysis can again be found in the OSF data analysis (https://osf.io/xqevy/files) and Supplementary Materials.

Same-sex ruling

Table 3.35S. Linear regression 1-(reciprocal+1) model: *Linear mixed model with four factors (positive language, negative language, number of followers, positive language x negative language) and number of retweets 1-(reciprocal+1) as the dependent variable.*

Fixed Effects					
	Estimate	*SE*	95% CI	*t*	*p*
Intercept	0.13	0.00052	0.13 - 0.14	261.04	.000
Positive Language Vader	-0.038	0.0014	-0.04 - -0.035	-27.46	.160
Negative Language Vader	-0.0078	0.0035	-0.018 - 0.00045	-2.19	.028
Number of followers	0.096	0.00034	0.096 - 0.097	282.64	.000
Positive Language V. × Negative Language V.	0.21	0.021	0.17- 0.25	9.98	.000
Random Effects					
			Variance	*SD*	
Participant (Intercept)			0.014	0.11	
Residual			0.034	0.19	
Model fit					
R^2			Marginal	Conditional	
			.16	.40	

Model equation: Retweets 1-(reciprocal+1) ~ Positive V. * Negative V. + centered (Followers) + (1 | User)

Notes. *Model fit was calculated using the R package MuMIn (Barton, 2018) based on the paper of Nakagawa et al. (2017).*

Table 3.36S. Model Comparison: *Comparing log-likelihood scores of SentiStrength vs. VADER models.*

	Marginal R^2	AIC	BIC	Log-Likelihood	df	chi-squared	*p*
SentiStrength		-139750	-139671	69847	7		
VADER	.16	140253	-140174	70106	7	88.12	.000

Notes. *Model fit was calculated using the R package lmtest (Hothorn et al., 2015)*

Ferguson unrest

Table 3.37S. Linear regression 1-(reciprocal+1) model: *Linear mixed model with four factors (positive language, negative language, number of followers, positive language × negative language) and number of retweets 1-(reciprocal+1) as the dependent variable.*

a.) Fixed Effects

	Estimate	*SE*	95% CI	*t*	*p*
Intercept	0.18	0.00067	0.18 - 0.19	276.92	.000
Positive Language Vader	-0.025	0.0034	-0.032 - -0.018	-7.45	.160
Negative Language Vader	0.010	0.0029	-0.0048 - 0.016	3.61	.000
Number of followers	0.099	0.00051	0.098 - 0.10	192.11	.000
Positive Language V. × Negative Language V.	0.26	0.028	0.20 - 0.32	9.19	.000

Random Effects

	Variance	SD
Participant (Intercept)	0.013	0.12
Residual	0.054	0.23

Model fit

R^2	Marginal	Conditional
	.11	.15

Model equation: Retweets 1-(reciprocal+1) ~ Positive V. * Negative V. + centered (Followers) + (1 | User)

Notes. *Model fit was calculated using the R package MuMIn (Barton, 2018) based on the paper of Nakagawa et al. (2017).*

Table 3.38S. Model Comparison: *Comparing log-likelihood scores of SentiStrength vs. VADER models.*

	Marginal R^2	AIC	BIC	Log-Likelihood	df	chi-squared	*p*
SentiStrength		40115	40193	-20084	7		
VADER	.16	40513	40591	-20275	7	382.17	.000

Notes. *Model fit was calculated using the R package lmtest (Hothorn et al., 2015)*

Analysis of most Viral Negative Tweets in Positive Contexts

I added the 10 most viral negative tweets to the supplementary material. I first categorized the tweets similarly to the topic model procedure, meaning the negative score has to be higher than the positive score. After that, we sorted them by number of retweets. The 10 most viral tweets in both situations were:

Trump election win

Table 3.39S. Most viral tweets: *Tweets with a negative sentiment score sorted by number of retweets.*

No.	Tweet content	Retweets
1.	They laughed at us when we said #realDonaldTrump would win. This morning they're too devastated to get out of bed	5759
2.	President elect Trump did what Obama never would he called families of slain officers to offer condolences	5606

3.	If you think they hate me now wait until I keep all my campaign promises #MAGA	4906
4.	I The People are awake amp we will expose amp vote out Establishment corruption #DrainTheSwamp #MAGA 2018	4315
5.	Soros Scrambling To Cover Up Evil Atheist Video From 1998 SEE IT BEFORE ITS DELETED #MAGA #AmericaFirst #Dobbs	4210
6.	America Is BURNING Property Is Being DESTROYED People are Being ATTACKED Barrys Silence Equals Consent #MAGA	4155
7.	Hillary conceded the election so that means that shes lied about absolutely everything this year #MAGA #AmericaFirst #realDonaldTrump	3201
8.	DonaldTrump didnt win today Hate won Fear won Racism won Sexism won Homophobia won Prejudice won	2536
9.	This scare me Trump #TheSimpsons #PresidentTrump #TrumpPresident #TrumpsFirstOrder LoPrimeroQueHaráTrump	2709
10.	A protest is an objection to a grievance If its just a bunch of sore losers. it's called a tantrum #MAGA	2709

Same-sex ruling

Table 3.40S. Most viral tweets: *Tweets with a negative sentiment score sorted by number of retweets.*

No.	Tweet content	Retweets
1.	today is honestly such a big day in history and we shouldnt forget how far we have come and all ppl have done to fight for equality #LoveWins	24017
2.	i cried #lovewins #happybirthdaytome #proudtobeanamerican af	17602
3.	before you decide to be an asshole just remember that there are people that have waited 50 years for today #LoveWins	2536
4.	On this day #LoveWins Gallavich Shameless Showtime	2457
5.	This church has a message for hateful members of their church #LoveWins	1433
6.	While today's decision is historic. We can't stop working until weve outlawed discrimination in every corner of our country #LoveWins HRC	1387
7.	Angry Americans saying they will flee to Canada hate to break it to you but #LoveWins there too	1342
8.	The biggest world news today were the attacks in Tunisia Kuwait and France The biggest news on twitter was #LoveWins of America	1193
9.	Fuck yeah #LoveWins	1035

10.	This ruling is a victory for gay and lesbian couples who have fought so long for their basic civil rights POTUS #LoveWins	879

I believe that this content mostly is congruent with the claim we made in the manuscript that negativity was mostly directed at the opposing group or the challenging path that the group has gone through to get to the celebrated moment. While there is an aspect of celebration in many of these tweets, there are also an abundance of negative emotions.

Removing Emotional Ambiguous Words from Same-sex Dataset

Similar to the initial topic modelling analysis, it seems that some of the negative tweets response to the same sex marriage are still expressing positive effect. One example is tweet No.2 Table 3.39S. ("I **cried** #lovewins #happybirthdaytome #proudtobeanamerican af"). Our second attempt to account for this limitation was to redo our analysis after removing all of the tweets who had the terms "cry", "sob", and "tear". After removing these tweets from the dataset we conducted the same analysis. Results shows similar effects to our main analysis, suggesting that the effect of these tweets on the over conclusion is marginal. I again used the same model as described above.

Same-sex ruling

Table 3.41S. Linear regression 1-(reciprocal+1) model: *Linear mixed model with four factors (positive language, negative language, number of followers, positive language x negative language) and number of retweets 1-(reciprocal+1) as the dependent variable.*

Fixed Effects					
	Estimate	*SE*	95% CI	*t*	*p*
Intercept	0.12	0.0016	0.13 - 0.14	72.56	.000
Positive Language Vader	0.00019	0.00076	-0.04 - -0.035	0.26	.794
Negative Language Vader	0.0086	0.0011	-0.018 - 0.00045	7.89	.000
Number of followers	0.094	0.00036	0.096 - 0.097	259.85	.000

Positive Language V. x Negative Language V.	0.00099	0.00051	0.17- 0.25	1.93	.052
Random Effects					
			Variance	*SD*	
Participant (Intercept)			0.020	0.14	
Residual			0.042	0.20	
Model fit					
R^2			Marginal	Conditional	
			.13	.41	

Model equation: Retweets 1-(reciprocal+1) ~ Positive V. * Negative V. + centered (Followers) + (1 | User)

Notes. *Model fit was calculated using the R package MuMIn (Barton, 2018) based on the paper of Nakagawa et al. (2017).*

CHAPTER 4: SUPPLEMENTARY MATERIALS

Linear Growth Models of Sentiment Change Over Time for Primary Social Sharing.

I re-conducted the analyses of the linear growth models to examine the relationship between users' levels of pandemic-specific primary social sharing and the sentiment scores of their COVID-related tweets throughout the crisis period. My approach adhered to the pre-registration document, and we employed a log + 1 transformation on the number of COVID-related original tweets, which served as a measure of the primary social sharing level, due to its positively skewed distribution. I again used multilevel modelling, as positive and negative sentiment scores were nested within days and participants. To evaluate the sentiment of these tweets, we utilized the VADER sentiment analysis tool and replicated my analysis as presented in the main manuscript, employing the same statistical models.

Results regarding the tweets' negative sentiment are reported in Table 4.1S and those regarding their positive sentiment are in Table 4.2S. An overview of the change in both sentiments is illustrated in Figure 4.1S.

Negative Sentiment Scores

Table 4.1S. *Linear Mixed Model with Five Factors (Primary Social Sharing Level, Time, Negative Sentiment During the Baseline Period, Overall Number of Tweets, Primary Social Sharing Level x Time)*

Fixed Effects					
	Estimate	*SE*	95% CI	*t*	*p*
Intercept	0.063	0.00036	0.062 – 0.064	174.68	.000
Primary Social Sharing Level	-0.0025	0.00036	-0.0032 – -0.0018	-6.80	.000
Time	-0.0012	0.00033	-0.0019 – -0.0006	-3.59	.000
Baseline Negative Sentiment	0.015	0.00024	0.015 – 0.016	65.95	.000
Overall Number of Tweets	0.0031	0.00028	0.0025 – 0.0036	10.59	.000
Primary Social Sharing Level × Time	-0.0007	0.00016	-0.0010 – -0.0004	-4.31	.000
Random Effects					
			Variance	*SD*	

Participant (Intercept)	0.00071	0.27
Time (Intercept)	0.0000017	0.0013
Residual	0.0071	0.084

Model Fit

R^2	Marginal	Conditional
	0.032	0.12

Model equation: Negative Sentiment Score ~ Primary Social Sharing Level * Time + Baseline Negative Sentiment + Overall Number of Tweets + (1 | User)

Note. *Model fit was calculated using the R package MuMIn (Barton & Barton, 2015) based on the paper of Nakagawa et al. (2017).*

Positive Sentiment Scores

Table 4.2S. *Linear Mixed Model with Five Factors (Primary Social Sharing Level, Time, Positive Sentiment During the Baseline Period, Overall Number of Tweets, Primary Social Sharing Level x Time)*

Fixed Effects

	Estimate	*SE*	95% CI	*t*	*p*
Intercept	0.10	0.00045	0.099 – 0.10	223.65	.000
Primary Social Sharing Level	0.00053	0.00045	-0.0030 – -0.0017	1.17	.000
Time	0.0023	0.00041	0.0010 – 0.0026	5.76	.000

Baseline Positive Sentiment	0.015	0.00028	0.0147 – 0.0158	53.49	.000
Overall Number of Tweets	-0.0034	0.00034	-0.0041 – -0.0028	-9.97	.000
Primary Social Sharing Level × Time	0.00066	0.00018	0.0003 – 0.0010	3.53	.000

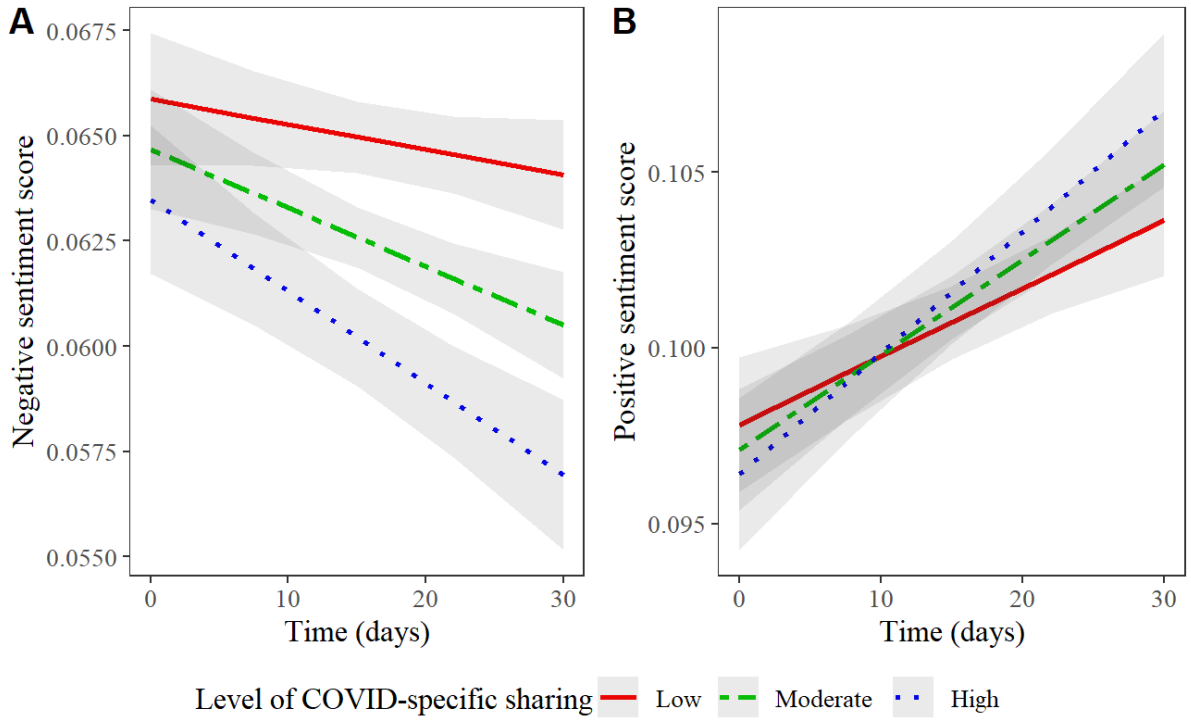
Random Effects		
	Variance	*SD*
Participant (Intercept)	0.0011	0.033
Time (Intercept)	0.0000025	0.0016
Residual	0.0097	0.098

Model Fit		
R^2	Marginal	Conditional
	0.022	0.12

Model equation: Positive Sentiment Score ~ Primary Social Sharing Level * Time + Baseline Positive Sentiment + Overall Number of Tweets + (1 | User)

Note. *Model fit was calculated using the R package MuMIn (Barton & Barton, 2015) based on the paper of Nakagawa et al. (2017).*

Figure 4.1S. *Sentiment Scores of COVID-Related Tweets as a Function of Time and Users' Level of Pandemic-Specific Primary Sharing*



Note. *Panel A shows negative sentiment scores as a function of time and users' level of pandemic-specific primary sharing while Panel B shows positive sentiment scores. The level of pandemic-specific primary sharing was a continuous variable but is presented as three categories in this figure to allow clearer visualization of the interaction effect. Different levels of sharing respectively modelled users who tweeted M – 1 SD times (minimum), the mean number of times (moderate), and M + 1 SD times (maximum).*

Linear Growth Models of Sentiment Change Over Time for Secondary Social Sharing.

Parallel to the analysis above, I re-conducted the analyses of the linear growth models to examine the relationship between users' levels of pandemic-specific secondary social sharing and the sentiment scores of their COVID-related tweets throughout the crisis period. Using the exact approach specified in the previous section.

Results regarding the tweets' negative sentiment are reported in Table 4.3S and those regarding their positive sentiment are in Table 4.4S. An overview of the change in both sentiments is illustrated in Figure 4.2S.

Negative Sentiment Scores

Table 4.3S. *Linear Mixed Model with Five Factors (Secondary Social Sharing Level, Time, Negative Sentiment During the Baseline Period, Overall Number of Tweets, Secondary Social Sharing Level x Time)*

Fixed Effects					
	Estimate	*SE*	95% CI	*t*	*p*
Intercept	0.063	0.00035	0.062 – 0.063	179.85	.000
Secondary Social Sharing Level	0.00010	0.00033	-0.0006 – 0.0008	0.30	.76
Time	-0.00079	0.00034	-0.0015 – -0.0001	-2.33	.019
Baseline Negative Sentiment	0.015	0.00023	0.015 – 0.016	65.93	.000
Overall Number of Tweets	0.0023	0.00028	0.0018 – 0.0029	8.36	.000
Secondary Social Sharing Level × Time	-0.00059	0.00016	-0.0009 – -0.0003	-3.77	.000
Random Effects					
			Variance	*SD*	
Participant (Intercept)			0.00071	0.27	
Time (Intercept)			0.0000017	0.0013	
Residual			0.0071	0.084	
Model Fit					
R^2			Marginal	Conditional	
			0.032	0.12	

Model equation: Negative Sentiment Score ~ Secondary Social Sharing Level * Time + Baseline Negative Sentiment + Overall Number of Tweets + (1 | User)

Notes. *Model fit was calculated using the R package MuMIn (Barton & Barton, 2015) based on the paper of Nakagawa et al. (2017).*

Positive Sentiment Scores

Table 4.4S. *Linear Mixed Model with Five Factors (Secondary Social Sharing Level, Time, Positive Sentiment During the Baseline Period, Overall Number of Tweets, Secondary Social Sharing Level x Time)*

Fixed Effects					
	Estimate	*SE*	95% CI	*t*	*p*
Intercept	0.10	0.00045	0.099 – 0.10	232.04	.000
Secondary Social Sharing Level	-0.0062	0.00045	-0.0070 – -0.0054	-14.95	.000
Time	0.0013	0.00041	0.0005 – 0.0021	3.08	.002
Baseline Positive Sentiment	0.015	0.00028	0.014 – 0.015	-5.04	.000
Overall Number of Tweets	-0.0017	0.00034	-0.0024 – -0.0010	53.01	.000
Secondary Social Sharing Level × Time	0.0003	0.00018	-0.0001 – 0.0007	1.59	.11

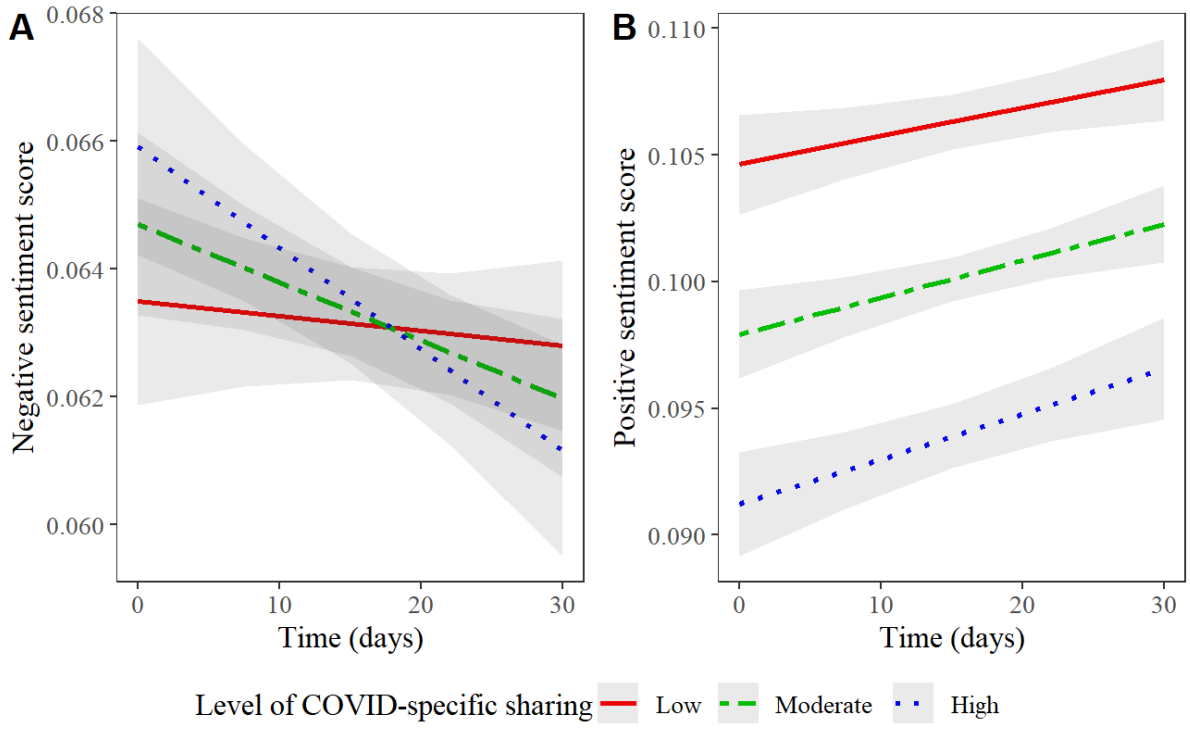
Random Effects		
	Variance	*SD*
Participant (Intercept)	0.0011	0.033

Time (Intercept)	0.0000025	0.0016
Residual	0.0097	0.098
Model Fit		
R^2	Marginal	Conditional
	0.025	0.13

Model equation: Positive Sentiment Score ~ Secondary Social Sharing Level * Time + Baseline Positive Sentiment + Overall Number of Tweets + (1 | User)

Notes. *Model fit was calculated using the R package MuMIn (Barton & Barton, 2015) based on the paper of Nakagawa et al. (2017).*

Figure 4.2S. *Sentiment Scores of COVID-Related Tweets as a Function of Time and Users' Level of Pandemic-Specific Secondary Sharing*



Note. *Panel A shows negative sentiment scores as a function of time and users' level of pandemic-specific secondary sharing while Panel B shows positive sentiment scores. The level of pandemic-specific secondary sharing was a continuous variable but is presented as three categories in this*

figure to allow clearer visualization of the interaction effect. Different levels of sharing respectively modelled users who tweeted M – 1 SD times (minimum), the mean number of times (moderate), and M + 1 SD times (maximum).

Dictionaries

Pre-defined Version (Short Dictionary)

I used a pre-registered dictionary for my analysis presented in the manuscript. This dictionary included terms that were most frequently used by participants in Study 2 when talking about COVID-related events according to my qualitative assessments: covid, ncov, lockdown, corona, social distance, virus, pandemic, covid19, flu, disease.

Extended Version (Long Dictionary)

I also adopted Dimitrov et al.'s (2020) TweetsCOV19 dataset, which included a more extensive list of terms, to re-select pandemic-related tweets to supplement my results. The terms were as follows: 14dayquarantine, anti-inflammation, anti-virus antibodies, antibody, antibody detection tests, antibody test, apron, atriumhealthproud, battleruncovid19, besafe, canceleverything, cdc, ce/fda, ce/fda certified, centre for disease control, cheer4healthworkers, china, china virus, chinavirus, chinese virus, chinesevirus, chloroquine, clean high touch surfaces regularly, cloth mask, convid19uk, corona, corona virus, coronakindness, coronapocalypse, coronavirus, coronavirus drug, coronavirus vaccine, coronaviruscrisis, coronavirusdeutschland, coronavirusimpact, coronavirusinsouthafrica, coronavirusisreal, coronavirusoutbreakindia, coronawarriors, coronials, covd, covid, covid tests, covid-19, covid-19 carrier, covid-patient, covid19, covid19 diagnosis, covid19 epidemic, covid19 nucleic acid test, covid19 pandemic, covid19ireland, covid19malaysia, covid19ph, covid19singapore, covid19uk, covid19update, covid_19, covidcarezw,

covidcoping, covidheroes, covidiot, covidiots, covidlife, covidresearch, covidwarriors, covid-19, decontamination, difficulty breathing, disinfectant, disinfection services, disinfection team, dontbeaspreader, don’t touch your face, dr. fraud fauci, duringmy14dayquarantine, ecmo, epidemic, epitwitter, essentialworkers, face masks, face shield, facemask, facemasks, ffp1, ffp2, ffp2 mask, ffp3, ffp3 mask, fightagainstcorona, fightagainstcoronavirus, fightcoronatogether, fightcovid19, fighttogether, flatten the curve, flattenthecurve, floridacoronavirus, frontline health care workers, getmeppe, gilead, goggle, hackthepandemic, hand-washing, health workers, healthcare, healthcareheroes, healthheroes, healthworkers, healthyathome, homequarantine, hydroxychloroquine, immune system, indiafightscorona, indiafightscoronavirus, indiafightscovid19, infection, infections, injectdisenfectant, inmyquarantinesurvivalkit, inthistogether, istayhomeforyou, kn95, kn95 mask, koronavirus, kungflu, lab4life, labheroes, lablife, labweek, labweek2020, level4lockdown, lock down, lockdown, lockdownextension, lockdownindia, lockdownzim, mask, masks4all, n-95 respirator masks, n95, n95 mask, n95 respirator mask, n95/kn95 respirator, ncov, nhscovidheroes, no shaking hands, nucleic acid, nucleic acid test, nucleic acid test kit, nucleic acid tests, nucleic-acid based test kits, nursescovid19, nypause, outbreak, pandamic, pandemia, pandemic, pandemic2020, pandemie, panic buy, panic shop, panic shopping, panic-buy, panic-shop, panicbuy, panicbuyingpanic buying, panicshop, personalprotectiveequipment, pharma, physicaldistancing, plankthecurve, plasma, ppe, ppeshortage, prayerforhealing, quarantine, quarantine shelter, quarantinelife, quarantineworkout, quarentinelife, remdesivir, respirator, safehands, saferathome, sanitizer, sars-cov-2, sars-cov-2 virus, sarscov2, savelives, saynobillgatesvaccine,sflockdown, sheltering in place, shelteringinplace, sinophobia, slow the spread of covid19, social distance, social distancing, socialdistance, socialdistancing, socialdistancingnow, standtogether,

stay at home, stay home, stay home challenge, stay safe stay home, stay_home_safe, stayathome, stayathomeandstaysafe, stayathomesa, stayathomesavelives, stayhealthy, stayhome, stayhome!,stayhomechallenge, stayhomesavelives, stayhomestaysafe, stayhometosavelives, stayinghome, stayingsafe, stayinstaysafe, staysafe, staysafestayhealthy, staysafestayhome, stopthespread, strongertogether, supportingnursesandmidwives, supportnursesandmidwives, supportsmallbusinesses, surgical gloves, surgical gwon, surgical mask, surgical nose mask, symptoms, tedros adhanom ghebreyesus, thankahealthcarehero, thankshealthheroes, thankyoucoronawarriors, thankyouessentialworkers, thankyouhealthcareworkers, togetherathome, togetherwecan, togetherwecandoit, treat coronavirus, trump pandemic, trumpdisinfectant, trumppandemic, uniteagainstcovid19, vaccination, vaccine, vaccine against coronavirus, vaccineswork, ventilator, videoconferencing, wash your hands, washyourhands, wear face mask, wearmask, wearmasks, wehealasone, who director-general, workfromhome, world health organization, wuhan, wuhancoronavirus, wuhanlockdown, yearofnurses, zimlockdown.

Validating the Performances of the Short and Long Dictionaries

To validate the performances of my two dictionaries, we randomly sampled 500 tweets from my pool of 1,683,322 tweets (regardless of their content) published by 29,484 users during the crisis period and asked two independent judges to code these tweets. The judges were asked to label each tweet as either 0 (non-COVID-related) or 1 (COVID-related). Interrater reliability was $r = .97$, meaning that the judges' codings differed for only five out of 500 tweets. Removing these five tweets, we compared the human-annotated classifications to those made by my two dictionaries.

The performance of each dictionary can be measured in terms of sensitivity and specificity. Sensitivity refers to the dictionary's ability to correctly classify tweets that human

judges determined were about the pandemic as COVID-related (i.e., true positive). The sensitivity level of the short dictionary was 87.61%, and that of the long dictionary was 92.92%. This suggests that the long dictionary was less subject to type II errors than the short dictionary. On the other hand, specificity refers to the dictionary's ability to correctly classify tweets that human judges determined were *not* about the pandemic as non-COVID-related (i.e., true negative). The specificity level of the short dictionary was 98.95%, and that of the long dictionary was 94.76%. This suggests that the short dictionary was less subject to type I errors than the long dictionary.

Overall, both dictionaries performed well in terms of sensitivity and specificity. I decided to focus on the results based on the tweets identified by the short dictionary in the manuscript for three reasons: 1) Assuming that there were more non-COVID-related tweets than COVID-related tweets in my pool, the ability to correctly reject non-COVID-related tweets would be more crucial in minimising misclassification. In this respect, the short dictionary was superior to the long one, 2) we pre-registered the short dictionary but not the long one, and 3) the two dictionaries did not yield significantly different results (comparisons will be presented in later sections).

Preregistered Pilot Study

Study 4's pilot study aimed to check the number of tweets (regardless of their content) ordinary users (excluding users who failed to indicate their location, 'verified' users and users who used COVID-related terms as part of their username) usually produced during the crisis period. Knowing this number was important to establish a threshold for exclusion, as users who produce considerably higher numbers of tweets than usual are more likely to be bots, which could significantly bias my data.

The first 10% of the data that we collected (number of tweets = 29,104) were analysed. The mean number of tweets each user produced was 91.18 and the range per user

was 1 - 600. The data distribution was right-skewed, indicating that there were several users who produced much higher numbers of tweets. Therefore, we believed that the mean, which is susceptible to outliers, was not an ideal measure of the central tendency in this dataset. Instead, we computed the median number of tweets users produced—it was 50.

I decided to set the threshold for exclusion at 200 tweets per month (+1 *SD* from the mean), meaning that users would only be considered as bots if they tweeted more than 4 times the median level. After implementing the threshold and filtering out possible bots, the mean number of tweets became 64.30, and the median number was 50.

Baseline Period

In addition to my main analysis, as presented in the manuscript, we collected the tweets produced by the same group of users between the 1st to 31st of October 2019 (1,740,368 tweets) to provide baseline measures that would allow comparisons between periods during the pandemic and before it.

Wang et al. (2022) tracked the evolution of expressed sentiment of social media posts in more than 100 countries between 1st January 2020 and 31st May 2020 and observed a steep decline in expressed sentiment globally immediately following the WHO's declaration. With their study period overlapping my chosen crisis period, we predicted that the general mood on Twitter should also be less positive in my dataset. I chose October 2019 as my baseline period because it was close enough in time to the first confirmed COVID case reported in both the UK and the US in January 2020. I avoided November and December, which were temporally closer to the official arrival of COVID-19 in these two countries because they are festive months (especially in the West) that generally elicit positive public emotions. For example, in their study, Bollen et al. (2011) found that on Thanksgiving Day in November, a traditional holiday mainly celebrated in the US and Canada, there was a sharp increase from baseline levels in the usage of terms that were classified as the emotion of vigour in users'

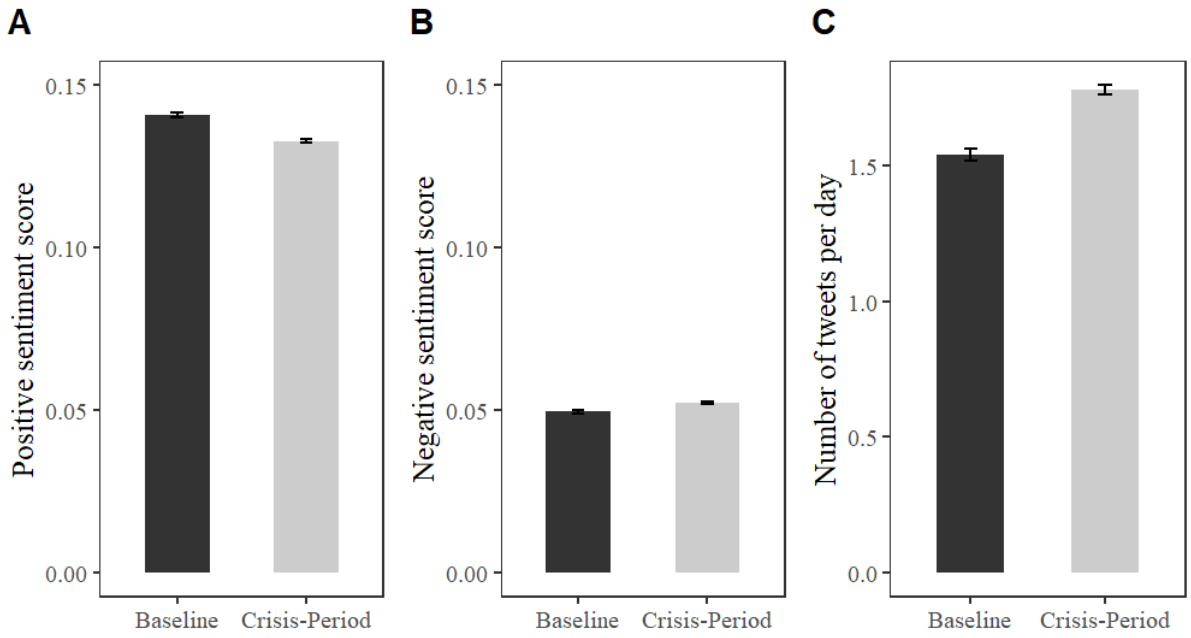
tweets. Similarly, Hu (2013) reported stronger positivity in users' tweets around Thanksgiving Day and even more so during the Christmas season in December. Furthermore, Twitter's API only allowed me to read the most recent 3,200 tweets produced by each user. At the time when we conducted the present studies, October was the earliest month in which we could collect most of the tweets produced by most of the users in my sample.

Comparing Sentiment Scores Between Baseline and Crisis Periods

I evaluated the assumption that the general mood online was negatively impacted by the pandemic during my chosen crisis period (20th April 2020 – 20th May 2020). I averaged sentiment scores (computed by VADER) of all kinds of tweets (COVID-related and not) published during the crisis period and the baseline period (1st – 31st October 2019) for comparison.

A paired t-test showed that positive sentiment scores were significantly lower ($t(29103) = -20.29$, $p < .001$; see Figure 4.3A) and negative sentiment scores were significantly higher ($t(29103) = 11.96$, $p < .001$; see Figure 4.3B) during the crisis period ($M_{pos} = 0.13$, $SD_{pos} = 0.053$; $M_{neg} = 0.049$, $SD_{neg} = 0.033$) than during the baseline period ($M_{pos} = 0.14$, $SD_{pos} = 0.067$; $M_{neg} = 0.052$, $SD_{neg} = 0.042$), consistent with my expectations. In addition, we found that there was a significant rise in the number of tweets per day during the crisis period ($t(29103) = 23.49$, $p < .001$; see Figure 4.3C), which indicated that the pandemic ($M = 1.78$, $SD = 1.82$) was associated with greater daily activity by Twitter users compared to the baseline period ($M = 1.53$, $SD = 2.03$). To summarise, users tweeted more and expressed less positive sentiment during the crisis period than before it.

Figure 4.3S *Sentiment Scores and Number of Tweets Posted During Baseline and Crisis Periods*



Notes. *Panels A and B respectively compare the average positive and negative sentiment scores of all kinds of tweets between the baseline (the left bar in every panel) and the crisis (the right bar) period. Panel C compares the number of all kinds of tweets per day between the two periods.*

Repeating VADER Analysis Using COVID-Related Tweets Identified by the Long Dictionary

I re-selected COVID-related tweets using the extended version of the dictionary. Then, we assessed the emotionality of those tweets using VADER and repeated my analysis presented in the main manuscript. The same statistical models were adopted. Results regarding the tweets' negative sentiment are reported in Table 4.5S. and those regarding their positive sentiment are in Table 4.6S. An overview of the change in both sentiments is illustrated in Figure 4.2S. I also compared the current results to the results based on the set of COVID-related tweets identified by the pre-defined dictionary in Tables 4.6S and 4.8S. In summary, the two sets of results were consistent with each other.

Negative Sentiment Scores

Table 4.5S *Linear Mixed Model with Five Factors (Social Sharing Level, Time, Negative Sentiment During the Baseline Period, Overall Number of Tweets, Social Sharing Level x Time)*

Fixed Effects					
	Estimate	*SE*	95% CI	*t*	*p*
Intercept	0.063	0.00036	0.062 – 0.064	174.68	.000
Social Sharing Level	-0.00013	0.00028	-0.00061 – 0.00041	-0.47	.63
Time	-0.0012	0.00036	-0.0019 – -0.00055	-3.38	.001
Baseline Negative Sentiment	0.015	0.00024	0.015 – 0.016	65.76	.000
Overall Number of Tweets	0.0030	0.00035	0.0023 – 0.0037	8.72	.000
Social Sharing Level × Time	-0.0010	0.00017	- 0.0014 – -0.00075	-5.98	.000
Random Effects					
			Variance	*SD*	
Participant (Intercept)			0.00071	0.27	
Time (Intercept)			0.0000017	0.0013	
Residual			0.0071	0.084	
Model Fit					
R^2			Marginal	Conditional	
			0.031	0.12	

Model equation: Negative Sentiment Score ~ Social Sharing Level * Time + Baseline Negative Sentiment + Overall Number of Tweets + (1 | User)

Notes. *Model fit was calculated using the R package MuMIn (Barton & Barton, 2015) based on the paper of Nakagawa et al. (2017).*

Table 4.6S *Comparison in Log-Likelihood Negative Sentiment Scores Between the Models Concerning the Two Different Sets of COVID-Related Tweets*

	R^2	AIC	BIC	Log-Likelihood	df	chi-squared	*p*
Short Dictionary	.12	-692604	-692507	346311	9		
Long Dictionary	.12	-692585	-692489	346302	9	18.33	.000

Notes. *Model fit was calculated using the R package lmtest (Hothorn et al., 2015).*

Positive Sentiment Scores

Table 4.7S *Linear Mixed Model with Five Factors (Social Sharing Level, Time, Positive Sentiment During the Baseline Period, Overall Number of Tweets, Social Sharing Level x Time)*

Fixed Effects					
	Estimate	*SE*	95% CI	*t*	*p*
Intercept	0.10	0.00044	0.099 – 0.10	223.65	.000
Social Sharing Level	-0.0023	0.00034	-0.0030 – -0.0017	-6.78	.000
Time	0.0018	0.00043	0.0010 – 0.0026	4.17	.000
Baseline Positive Sentiment	0.015	0.00028	0.014 – 0.016	52.70	.000

Overall Number of Tweets	-0.0026	0.00042	-0.0034 – -0.0017	-6.20	.000
Social Sharing Level × Time	0.00095	0.00021	0.00045 – 0.0013	4.51	.000
Random Effects					
			Variance	*SD*	
Participant (Intercept)			0.0011	0.033	
Time (Intercept)			0.0000025	0.0016	
Residual			0.0097	0.098	
Model Fit					
R^2			Marginal	Conditional	
			0.025	0.13	

Model equation: Positive Sentiment Score ~ Social Sharing Level * Time + Baseline Positive Sentiment + Overall Number of Tweets + (1 | User)

Notes. *Model fit was calculated using the R package MuMIn (Barton & Barton, 2015) based on the paper of Nakagawa et al. (2017).*

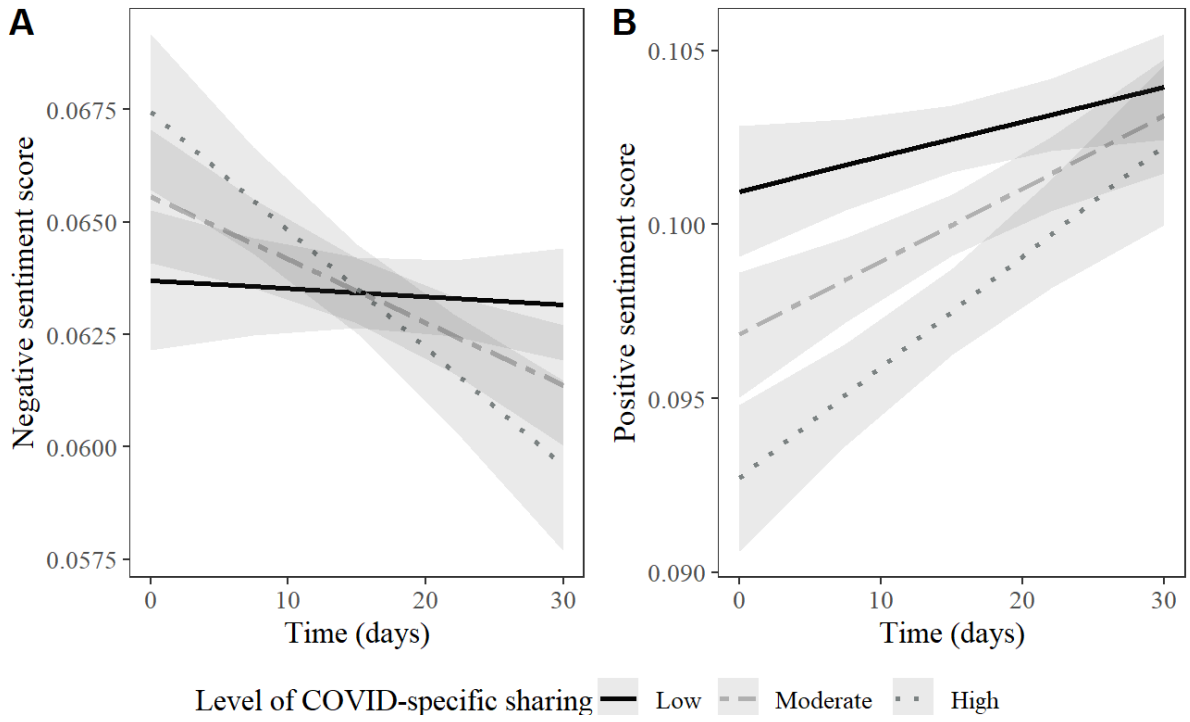
Table 4.8S *Comparison in Log-Likelihood Positive Sentiment Scores Between the Models Concerning the Two Different Sets of COVID-Related Tweets*

	R^2	AIC	BIC	Log-Likelihood	df	chi-squared	*p*
Short Dictionary	.13	-583280	-583183	291649	9		
Long Dictionary	.13	-583312	-583215	291665	9	32.32	.000

Notes. *Model fit was calculated using the R package lmtest (Hothorn et al., 2015).*

Overview of Change in Sentiment Scores of COVID-Related Tweets Identified By the Long Dictionary

Figure 4.4S *Sentiment Scores of COVID-Related Tweets as a Function of Time and Users' Level of Pandemic-Specific Sharing*



Notes. *Panel A shows negative sentiment scores as a function of time and users' level of pandemic-specific sharing while Panel B shows positive sentiment scores. The level of pandemic-specific sharing was a continuous variable but is presented as three categories in this figure to allow clearer visualization of the interaction effect. Different levels of sharing respectively modelled users who tweeted M – 1 SD times (minimum), the mean number of times (moderate), and M + 1 SD times (maximum).*

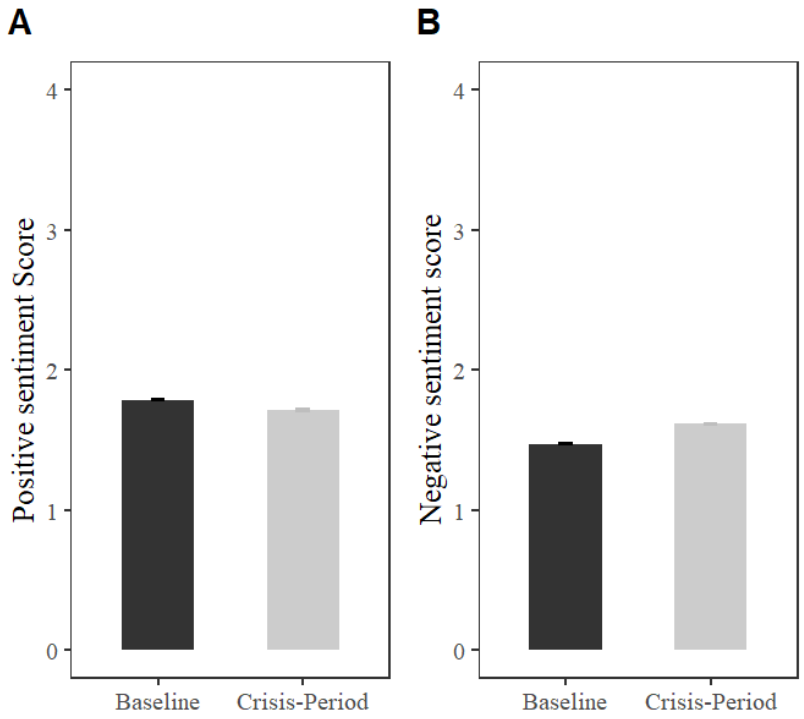
Repeating All Analyses Concerning COVID-Related Tweets Identified by the Short Dictionary Using SentiStrength

I re-conducted all my analyses concerning COVID-related tweets identified by the pre-defined dictionary using another sentiment analysis tool, SentiStrength (Thelwall et al., 2010). As can be seen from the details presented below, the results were consistent with those we obtained using VADER.

Comparing Sentiment Scores Between Baseline and Crisis Periods

A paired t-test showed that positive sentiment scores were significantly lower ($t(29103) = 31.44$, $p < .001$; see Figure 4.5A) and negative sentiment scores were significantly higher ($t(29103) = 12.31$, $p < .001$; see Figure 4.5B) during the crisis period ($M_{pos} = 1.78$, $SD_{pos} = 0.48$; $M_{neg} = 1.61$, $SD_{neg} = 0.53$) than during the baseline period ($M_{pos} = 1.71$, $SD_{pos} = 0.45$; $M_{neg} = 1.47$, $SD_{neg} = 0.38$).

Figure 4.5S Sentiment Scores During Baseline and Crisis Periods



Notes. *Panels A and B respectively compare the average positive and negative sentiment scores of all kinds of tweets between the baseline (the left bar in every panel) and the crisis (the right bar) period.*

Change in Sentiment Scores of COVID-Related Tweets

Next, we assessed the temporal change in positive and negative sentiment scores of COVID-related tweets for participants who engaged in different levels of social sharing. The statistical models were the same as in the main manuscript. SentiStrength results regarding the tweets' negative sentiment are reported in Table 4.9S, and those regarding their positive sentiment in Table 4.11S. Figure 4.6S shows an overview of the change in both sentiments over time. Tables 4.10S. and 4.12S. compare the current results to my original VADER results. In summary, we found that VADER seemed to be a more suitable tool than SentiStrength for sentiment analysis, as my model comparisons showed that the former offered a better model fit.

Table 4.9S *Linear Mixed Model with Five Factors (Social Sharing Level, Time, Negative Sentiment During the Baseline Period, Overall Number of Tweets, Social Sharing Level x Time)*

Fixed Effects					
	Estimate	*SE*	95% CI	*t*	*p*
Intercept	1.75	0.0037	1.74 – 1.76	467.29	.000
Social Sharing Level	-0.012	0.0029	-0.018 – -0.0050	-4.04	.000
Time	-0.017	0.0035	-0.024 – -0.010	-5.12	.000
Baseline Negative Sentiment	0.17	0.0027	0.16 – 0.18	63.04	.000
Overall Number of Tweets	0.027	0.0034	0.019 – 0.034	7.82	.000
Social Sharing Level × Time	-0.0089	0.0018	-0.012 – -0.0047	-4.91	.000
Random Effects					
			Variance	*SD*	
Participant (Intercept)			0.093	0.30	
Time (Intercept)			0.00013	0.011	
Residual			0.82	0.90	
Model Fit					
R^2			Marginal	Conditional	
			0.031	0.13	

Model equation: Negative Sentiment Score ~ Social Sharing Level * Time + Baseline Negative Sentiment + Overall Number of Tweets + (1 | User)

Notes. *Model fit was calculated using the R package MuMIn (Barton & Barton, 2015) based on the paper of Nakagawa et al. (2017).*

Table 4.10S *Comparison in Log-Likelihood Negative Sentiment Scores Between the Models Using Two Different Sentiment Analysis Tools*

	R^2	AIC	BIC	Log-Likelihood	df	chi-squared	*p*
SentiStrength	.11	913386	913483	-456684	9		
VADER	.13	-692604	-692507	346311	9	1605990	.000

Notes. *Model fit was calculated using the R package lmtest (Hothorn et al., 2015).*

Table 4.11S *Linear Mixed Model with Five Factors (Social Sharing Level, Time, Positive Sentiment During the Baseline Period, Overall Number of Tweets, Social Sharing Level x Time)*

Fixed Effects					
	Estimate	*SE*	95% CI	*t*	*p*
Intercept	1.60	0.0038	1.59 – 1.61	467.29	.000
Social Sharing Level	-0.012	0.0025	-0.041 – -0.030	-14.03	.000
Time	-0.017	0.0035	-0.015 – 0.0027	-1.86	.068
Baseline Positive Sentiment	0.14	0.0022	0.13 – 0.14	62.70	.000
Overall Number of Tweets	0.0050	0.0029	-0.00021 – 0.011	1.69	.089
Social Sharing Level × Time	0.0017	0.0015	-0.0012 – 0.0050	1.12	.26
Random Effects					
			Variance	*SD*	
Participant (Intercept)			0.069	0.26	

Time (Intercept)	0.00023	0.015
Residual	0.60	0.77
Model Fit		
R^2	Marginal	Conditional
	0.031	0.13

Model equation: Positive Sentiment Score ~ Social Sharing Level * Time + Baseline Positive Sentiment + Overall Number of Tweets + (1 | User)

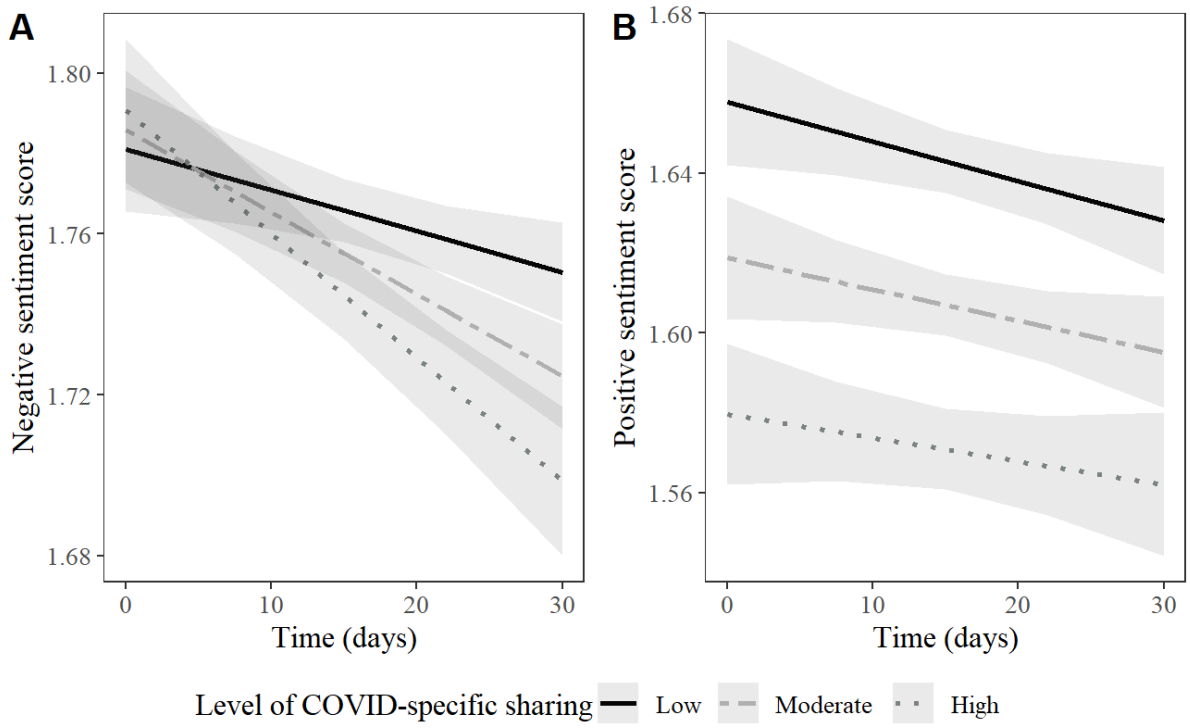
Notes. *Model fit was calculated using the R package MuMIn (Barton & Barton, 2015) based on the paper of Nakagawa et al. (2017).*

Table 4.12S *Comparison in Log-Likelihood Positive Sentiment Scores Between the Models Using Two Different Sentiment Analysis Tools*

	R^2	AIC	BIC	Log-Likelihood	df	chi-squared	*p*
SentiStrength	.12	913386	913483	-456684	9		
VADER	.13	-692604	-692507	346311	9	1605990	.000

Notes. *Model fit was calculated using the R package lmtest (Hothorn et al., 2015).*

Figure 4.6S *Sentiment Scores of COVID-Related Tweets as a Function of Time and Users' Level of Pandemic-Specific Sharing*



Notes. *Panel A shows negative sentiment scores as a function of time and users' level of pandemic-specific sharing while Panel B shows positive sentiment scores. The level of pandemic-specific sharing was a continuous variable but is presented as three categories in this figure to allow clearer visualization of the interaction effect. Different levels of sharing respectively modelled users who tweeted M – 1 SD times (minimum), the mean number of times (moderate), and M + 1 SD times (maximum).*

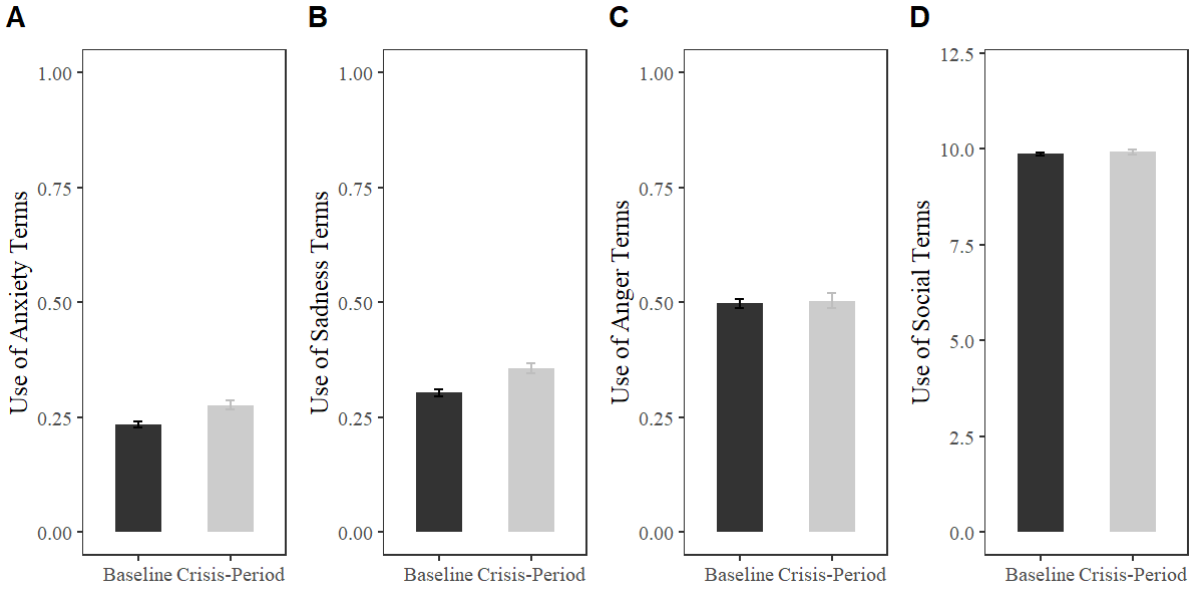
Preregistered LIWC Analyses

I re-processed all my tweets using LIWC (Pennebaker et al., 2015) as preregistered. Specifically, we used the LIWC categories of "anxiety," "sadness," and "anger," to estimate the negativity of tweets and "social" (which contains terms that refer to other people, relationships, and social activities) to estimate the positivity of tweets.

Comparing Usage of Terms Between Baseline and Crisis Periods

A paired t-test showed that the use of both anxiety terms ($t(29103) = 6.98$, $p < .001$; see Figure 4.7A) and sadness terms ($t(29103) = 8.38$, $p < .001$; see Figure 4.7B) was significantly higher during the crisis period than during the baseline period. The use of anger terms ($t(29103) = 0.60$, $p = .54$; see Figure 4.7C) and social terms ($t(29103) = 1.59$, $p = .11$; see Figure 4.7D), however, was not significantly different between the two measurement periods.

Figure 4.7S *Usage of Terms During Baseline and Crisis Periods*



Notes. *Panels A, B, C, and D respectively compare the average use of anxiety, sadness, anger, and social terms of all kinds of tweets between the baseline (the left bar in every panel) and the crisis (the right bar) period.*

Change in Usage of Terms in COVID-Related Tweets Identified by the Short Dictionary

Next, we focused on COVID-related tweets identified by the pre-defined dictionary and assessed the use of anxiety, sadness, anger, and social terms in these tweets over time for

users who engaged in different levels of social sharing. The structure of the statistical models used here was the same as in the main manuscript.

I found that users who tweeted about the pandemic more often tended to use fewer sadness and anger terms than those who tweeted less about the same topic, consistent with my VADER and SentiStrength finding that higher levels of pandemic-specific sharing were associated with a reduction in negative sentiment scores. However, the LIWC results did not show a significant interaction between pandemic-specific sharing and time on the use of anxiety or social terms. This pattern of results suggests that not all kinds of negative sentiments were equally affected by COVID-related sharing (e.g., the effects were stronger for sadness and anger than for anxiety).

Table 4.13S *Linear Mixed Model with Five Factors (Social Sharing Level, Time, Anxiety Terms During the Baseline Period, Overall Number of Tweets, Social Sharing Level x Time)*

Fixed Effects					
	Estimate	*SE*	95% CI	*t*	*p*
Intercept	0.30	0.0040	0.29 – 0.31	75.53	.000
Social Sharing Level	0.034	0.0039	0.027 – 0.042	8.60	.000
Time	0.0056	0.0045	-0.0021– 0.013	1.25	.21
Baseline Anxiety Terms	0.067	0.0033	0.061– 0.074	20.48	.000
Overall Number of Tweets	-0.016	0.0049	-0.024 – -0.0064	-3.30	.000
Social Sharing Level × Time	-0.00035	0.0029	-0.0069 – 0.0045	-0.12	.90

Random Effects		
	Variance	*SD*
Participant (Intercept)	0.082	0.28
Time (Intercept)	0.000098	0.0099
Residual	2.32	1.52
Model Fit		
R^2	Marginal	Conditional
	0.0023	0.036

Model equation: Anxiety Terms ~ Social Sharing Level * Time + Baseline Anxiety Terms + Overall Number of Tweets + (1 | User)

Notes. *Model fit was calculated using the R package MuMIn (Barton & Barton, 2015) based on the paper of Nakagawa et al. (2017).*

Table 4.14S *Linear Mixed Model with Five Factors (Social Sharing Level, Time, Sadness Terms During the Baseline Period, Overall Number of Tweets, Social Sharing Level x Time)*

Fixed Effects					
	Estimate	*SE*	95% CI	*t*	*p*
Intercept	0.33	0.0054	0.32 – 0.34	61.54	.000
Social Sharing Level	-0.026	0.0046	-0.036 – -0.016	-5.67	.000
Time	-0.013	0.0059	-0.022 – -0.0021	-2.21	.031
Baseline Sadness Terms	0.026	0.0057	0.012 – 0.039	4.65	.000
Overall Number of Tweets	0.047	0.0038	0.038 – 0.056	12.51	.089
Social Sharing Level × Time	-0.0066	0.0033	-0.012 – 0.00080	-1.96	.049
Random Effects					

	Variance	*SD*
Participant (Intercept)	0.10	0.32
Time (Intercept)	0.00036	0.018
Residual	3.13	1.76

Model Fit

R^2	Marginal	Conditional
	0.00089	0.034

Model equation: Sadness Terms ~ Social Sharing Level * Time + Baseline Sadness Terms + Overall Number of Tweets + (1 | User)

Notes. *Model fit was calculated using the R package MuMIn (Barton & Barton, 2015) based on the paper of Nakagawa et al. (2017).*

Table 4.15S *Linear Mixed Model with Five Factors (Social Sharing Level, Time, Anger Terms During the Baseline Period, Overall Number of Tweets, Social Sharing Level x Time)*

Fixed Effects

	Estimate	*SE*	95% CI	*t*	*p*
Intercept	0.45	0.0081	0.32 – 0.34	55.62	.000
Social Sharing Level	-0.067	0.0057	-0.036 – -0.016	-11.73	.000
Time	-0.031	0.0084	-0.022 – -0.0021	-3.68	.000
Baseline Anger Terms	0.067	0.0071	0.012 – 0.039	9.39	.000
Overall Number of Tweets	0.19	0.0047	0.038 – 0.056	39.72	.000
Social Sharing Level × Time	-0.014	0.0040	-0.013 – -0.0010	-3.50	.000

Random Effects

	Variance	*SD*

Participant (Intercept)	0.20	0.45
Time (Intercept)	0.0010	0.032
Residual	4.46	2.11

Model Fit

R^2	Marginal	Conditional
	0.00089	0.052

Model equation: Anger Terms ~ Social Sharing Level * Time + Baseline Sadness Terms + Overall Number of Tweets + (1 | User)

Notes. *Model fit was calculated using the R package MuMIn (Barton & Barton, 2015) based on the paper of Nakagawa et al. (2017).*

Table 4.16S *Linear Mixed Model with Five Factors (Social Sharing Level, Time, Social Terms During the Baseline Period, Overall Number of Tweets, Social Sharing Level x Time)*

Fixed Effects

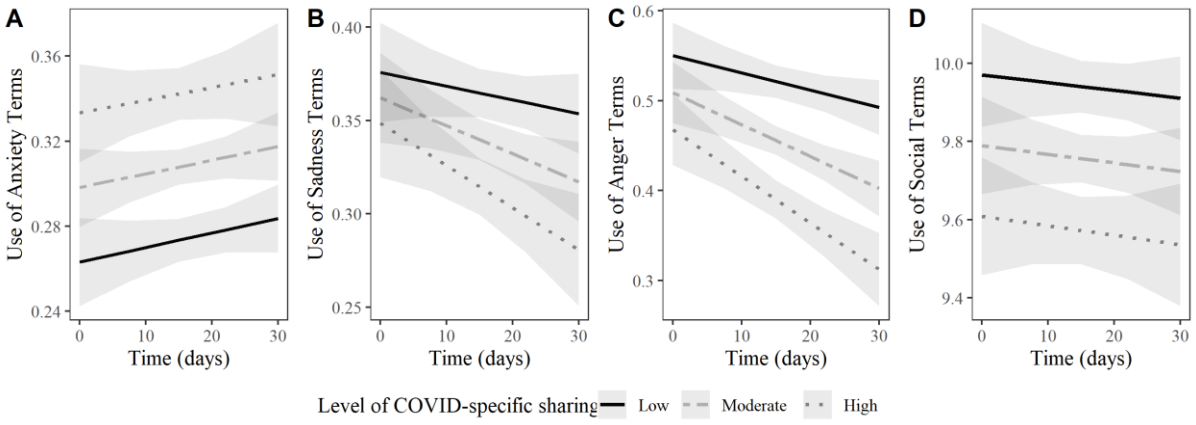
	Estimate	*SE*	95% CI	*t*	*p*
Intercept	0.33	0.030	0.32– 0.34	317.77	.000
Social Sharing Level	-0.18	0.024	-0.035 – -0.015	-7.52	.000
Time	-0.019	0.030	-0.026 – -0.000028	-0.64	.52
Baseline Social Terms	0.067	0.029	0.014 – 0.037	2.29	.021
Overall Number of Tweets	1.39	0.021	0.039 – 0.054	65.58	.000
Social Sharing Level × Time	-0.0019	0.015	-0.014 – -0.000086	-0.12	.90

Random Effects

	Variance	*SD*
Participant (Intercept)	5.25	2.29
Time (Intercept)	0.010	0.10
Residual	64.95	8.05
Model Fit		
R^2	Marginal	Conditional
	0.00089	0.052
Model equation: Social Terms ~ Social Sharing Level * Time + Baseline Social Terms + Overall Number of Tweets + (1 | User)		

Note. *Model fit was calculated using the R package MuMIn (Barton & Barton, 2015) based on the paper of Nakagawa et al. (2017).*

Figure 4.8S *Usage of Terms in COVID-Related Tweets as a Function of Time and Users' Level of Pandemic-Specific Sharing*



Note. *Panels A, B, C, and D respectively show the use of anxiety, sadness, anger, and social terms in COVID-related tweets as a function of time and users' level of pandemic-specific sharing. The level of sharing was a continuous variable but is presented as three categories in this figure to allow clearer visualization of the interaction effect. Different levels of sharing respectively modelled users who*

tweeted M – 1 SD times (minimum), the mean number of times (moderate), and M + 1 SD times (maximum).

Repeating the Granger-Causality Test Using 2 Weeks Instead of 1 Week as the Time Period

I repeated the Granger-Causality tests by setting the time period to 2 weeks and the time lag as 1, meaning that we aggregated my data across two-week periods and assessed the extent to which the number of COVID-related tweets in the first two-week period ($T - 1$) predicted sentiment scores for the COVID-related tweets from the following two-week period (T).

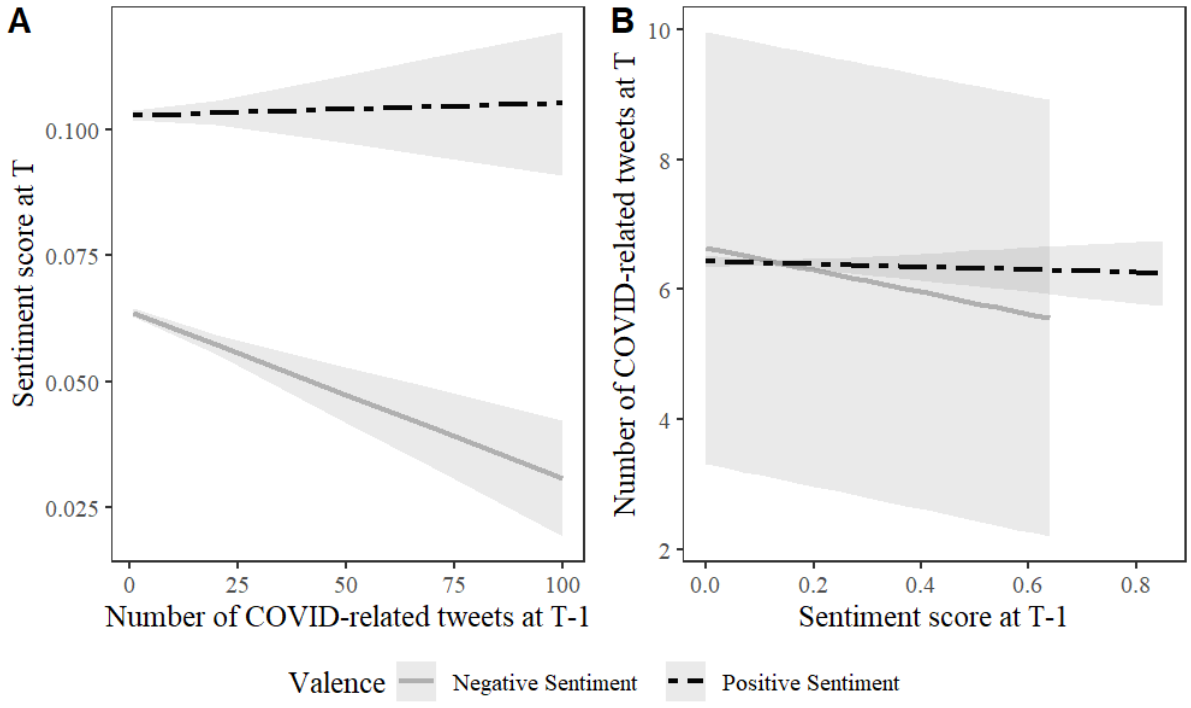
As described in the main manuscript, the Granger-Causality test involves a simple and a complex model, so we compared the two models using data aggregated from two-week instead of one-week periods for positive and negative sentiment scores separately.

In the case of negative sentiment scores, we found that the complex model had a better fit than the simple model ($\chi^2 = 28.57$, $p < .001$), indicating that a higher level of pandemic-specific sharing during the first two weeks of the study 'Granger-caused' a decrease in the use of negative-affect words in COVID-related tweets in the following two weeks ($b = -0.00033$, $SE = 0.000062$, $t(28984.13) = -5.34$, $p < .001$, 95% CI [-0.00045, -0.00021]; see Figure 4.9A). Running the test in the reverse direction, we again found that the complex model provided a better fit than the simple model ($\chi^2 = 4172.1$, $p < .001$), suggesting that increased use of negative-affect words in COVID-related tweets in the first two weeks of the study predicted a decrease in pandemic-specific sharing levels in the following two weeks ($b = -0.11$, $SE = 0.025$, $t(37679.00) = -4.17$, $p < .001$, 95% CI [-0.18, -0.055]; see Figure 4.9B). I then compared the fits of both directional models and found that the forward-

causation model (*AIC* = -109903) had a better fit than the reverse-causation model (*AIC* = 224421). In other words, the influence of prior social sharing on subsequent negativity was greater than that of prior negativity on subsequent social sharing.

In the case of positive sentiment scores, the Granger-Causality test again did not yield any significant results for forward causation (p = .52) or reverse causation (p = .77). To summarise, all the results obtained when the time period was set to 2 weeks were consistent with those presented in the manuscript (time period as 1 week).

Figure 4.9S *Temporal Causality Between Sentiment Scores and Social Sharing of COVID-Related Content*



Note. *Panel A illustrates how pandemic-specific sharing frequency in the last two weeks predicted sentiment scores of COVID-related tweets posted during the next two weeks (i.e., forward causation). Panel B, on the other hand, illustrates how sentiment scores of COVID-related tweets posted during the last two weeks predicted pandemic-specific sharing frequency in the next two weeks (i.e., reverse*

causation). T – 1 stand for a time period (2 weeks) going back by a time lag of 1—that is, two weeks before. The darker lines represent positive sentiment scores of COVID-related tweets and the lighter lines negative sentiment scores.

CHAPTER 5: SUPPLEMENTARY MATERIALS

Study 9: Evaluation of Affect in Word Sequences

The goal of Study 9 was to test whether people are able to accurately evaluate the average affect expressed in sequences of words. In my pre-registration (https://osf.io/a4v3d), I hypothesized that: (1) participants would overestimate the emotional language intensity of the sequence compared to the computed average based on the individual ratings, (2) this overestimation would increase for longer sequences, and (3) negative sequences would elicit greater overestimation compared to positive ones.

Method

Participants. To ensure that participants were familiar with English words and could accurately judge their emotional valence, I limited our recruitment to individuals residing in the United Kingdom or the United States. I recruited 100 participants via the Prolific participant provider platform (66 females, 33 males, 1 other, age: M = 41.86, SD = 15.39, range = 18-74). All participants in each of the following studies provided informed consent and were financially reimbursed. Specifically, in this experiment they were compensated $2.30 for their participation in the 20-minute task. All studies were approved by the Human

Subject Committee at Harvard University (IRB20-0091) and University of Oxford's Medical Sciences Inter-Divisional Research Ethics Committee (IDREC, R74339).

Stimulus set. I obtained the words for sequences in this experiment from the Evaluation Lexicon 2.0 (EL 2.0), developed by Rocklage, Rucker, and Nordgren (2018), which is a sentiment analysis tool designed to measure emotionality, extremity, and valence in language. Each of the 1,541 words in this dictionary was rated approximately 30 times on each of these dimensions. These ratings were used to calculate the objective emotionality of a sequence in our analysis. To avoid repetitions, I only included word stems in our stimulus set, omitting all word derivatives. For example, both "happily" and "happy" existed in the EL 2.0, but I chose to retain only the word stem "happy" in the final stimulus pool. Additionally, I only included words that were in both the EL 2.0 and the ANEW dictionary compiled by Bradley and Lang (1999), because the ANEW dictionary provides a rating for a word's frequency of use in everyday language, which enabled me to distinguish between a word's frequency and its intensity rating in my analysis These steps narrowed down my selection to 853 words that were presented to participants in the task.

Procedure. Each participant completed 50 trials preceded by 2 practice trials, which were excluded from the analysis. At the start of each trial, participants were shown a sequence of 1-15 words expressing varying degrees of affect from either a neutral-to-negative (negative condition) or a neutral-to-positive (positive condition) continuum. Each word in the sequence was randomly selected from a list of 853 stimuli and displayed on a white background in the centre of the screen for 1000ms. No participant saw the same word more than once. Between words, participants saw a fixation cross for a duration between 400-600ms (randomly determined on each trial). The valence and the number of words presented in each sequence were also chosen randomly in each trial. I chose not to mix positive and negative words in each sequence to avoid potentially undermining our ability to detect an

amplification effect. Specifically, if participants' judgments were equally affected by both extremely negative and extremely positive words, their final estimate of the sequence's average emotional valence might appear to be accurate, even if their perceptions were disproportionately influenced by the more emotionally intense words of each valence that they encountered. After each sequence, participants were asked to "*estimate the average emotion expressed by the sequence of words*" on a scale from 0 (Not at all Emotional) to 9 (Very Emotional). To estimate the mean intensity of the emotional sequence, participants were given as much time as they needed. After completing the main task, participants filled out a short survey including demographic data and a personality as well as social anxiety scale (for full detail, see SI).

Measures. In order to measure amplification, I compared the estimated mean with the actual mean affect of the sequence. The actual sequence mean was calculated by taking the mean affect displayed by each word, as rated on the emotionality dimension of the EL 2.0. Both ratings used the same 0 ("Not at all Emotional") to 9 (indicating "Very Emotional") scale.

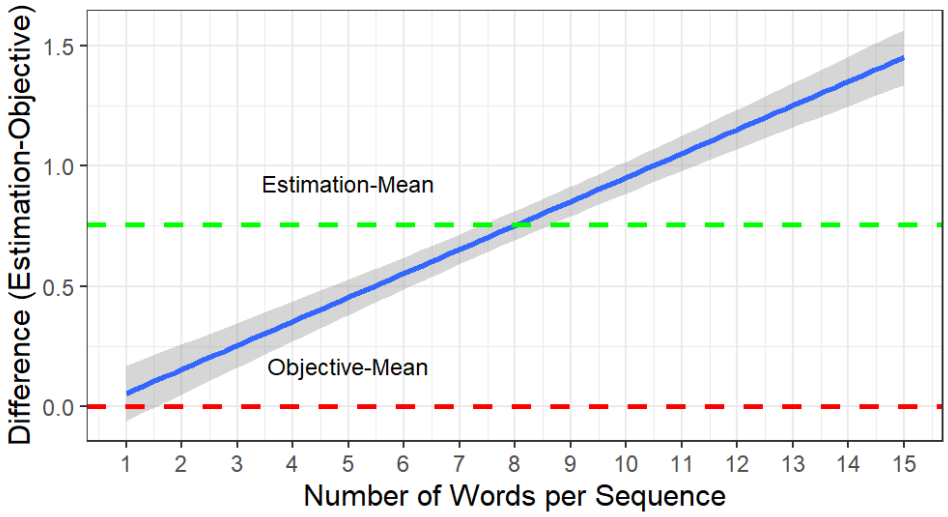
Results

The analysis and packages used in the current experiment were identical to those utilized in Studies 5 - 9, with two exceptions. First, I added the frequency of a word in everyday language, which was extracted from the ANEW dictionary, as a covariate to ensure that the result was not due to participants focusing on relatively rarer words. I also included random intercepts for participants to adjust for individual differences in emotional language perception for each model. Second, as there was no news article in this experiment, the news article rating was excluded as a covariate.

Overestimation of the average emotional language expressed. To evaluate whether participants overestimate the average affect expressed in a sequence of words, I compared the difference in emotional expression values between two types of means: first, the actual sequence mean evaluated in the EL 2.0, and second, the estimated mean of the sequence as reported by the participant, in our analysis. Supporting the first hypothesis, the estimated mean emotional language was higher than the actual mean emotional language (b = 0.75 [0.69, 0.82], SE = 0.032, t (9893.93) = 23.53, $p < .001$, $R^2 = .15$).

Overestimation of the average emotional language expressed by sequence length. To test my second hypothesis, I investigated whether the length of the sequence affected participants' estimation of the mean. As in experiments 1 - 2, the number of responses per sequence was the independent variable, but in this case, the sequence length ranged from 1 to 15, while keeping the emotionality rating of the news article constant. I subtracted one from the sequence length, so that the intercept represents the estimation difference when the sequence consists of only one word. I found that an increase in the number of words per sequence was associated with a stronger overestimation of the mean (b = 0.75 [0.69, 0.82], SE = 0.032, t (9893.93) = 23.53, $p < .001$, $R^2 = .30$, see Figure 5.1S). Supporting the claim that the estimation of an individual word was accurate, when a sequence consisted of only one word (represented by the intercept of the model), the difference between the objective mean and the estimation was not significantly different from zero (b = -0.17 [0.69, 0.82], SE = 0.19, t (670.40) = -0.92, $p = .36$, $R^2 = .30$).

Figure 5.1S *The comparison between estimated and actual average ratings of the sequence. Analysis was done using a linear mixed model.*



Notes. *The x-axis represents the number of words in the sequence. The y-axis represents the difference between participants' estimation of the average sequence and the actual average. Data are presented as mean value +/- confidence intervals. The green dotted line represents the average amplification. The red dotted line represents the actual emotional intensity of the word's mean.*

Overestimation of the average emotion expressed by valence. Finally, I wanted to assess whether participants' overestimations differed between positive and negative sequences. I found that participants significantly overestimated the mean more for negative compared to positive sequences (b = 0.54 [0.44, 0.65], SE = 0.053, t (4905.18) = 10.12, p < .001, R^2 = .29).

Discussion

In Experiment 5, I found evidence that people tended to overestimate the affect expressed in sequences of words. I also observed that this overestimation was more pronounced for longer sequences and stronger for negative compared to positive sequences.

Despite the similarities between the findings of the studies using sentences to this study that uses words, it is unclear whether the mechanisms leading to both findings are the same. Specifically, I argue that one of the reasons for amplifications in sentences is that people perceive more emotionally intense sentences as more representative for the sequence. However, when it comes to word sequences, the concept of representativeness is unlikely to play a significant role since there is no contextual information (such as the news article in the sentence study) for a word to be representative of. It seems more plausible that the amplification of words is similar to the amplification of faces, as both stimuli lack contextual or deeper meaning behind their emotional impact. In contrast, sentences involve not only emotional content but also convey semantic meaning that requires processing. Nevertheless, it is important to note that there is currently no empirical evidence available to support this claim. Future studies could investigate the mechanisms of amplifications in words and whether the amplification in words contributes to the overestimation of sentences.

One limitation of this study is that I used the ratings of previous rater to determine the objective mean, and similar to the study by Brady et al. (2022), potential differences between the previous raters and the current participants could lead to the amplification effect. Despite this limitation, the findings from this experiment provide suggestive evidence that the amplification of emotional stimuli may extend to other domains. For example, future research could explore the amplification phenomenon in relation to sequences involving the aggregation of rewards or punishments, broadening our understanding of the generalizability of the amplification effect.

Power Calculations to Estimate Experiment Sample Sizes

Experiment 5

The sample size of this study was evaluated by a triangulation of a few goals, the first goals was designed to achieve enough responses to be able to receive enough evaluations for each response, which in turn allowed us to screen some of the existing responses. The second goal was to detect the amplification effect.

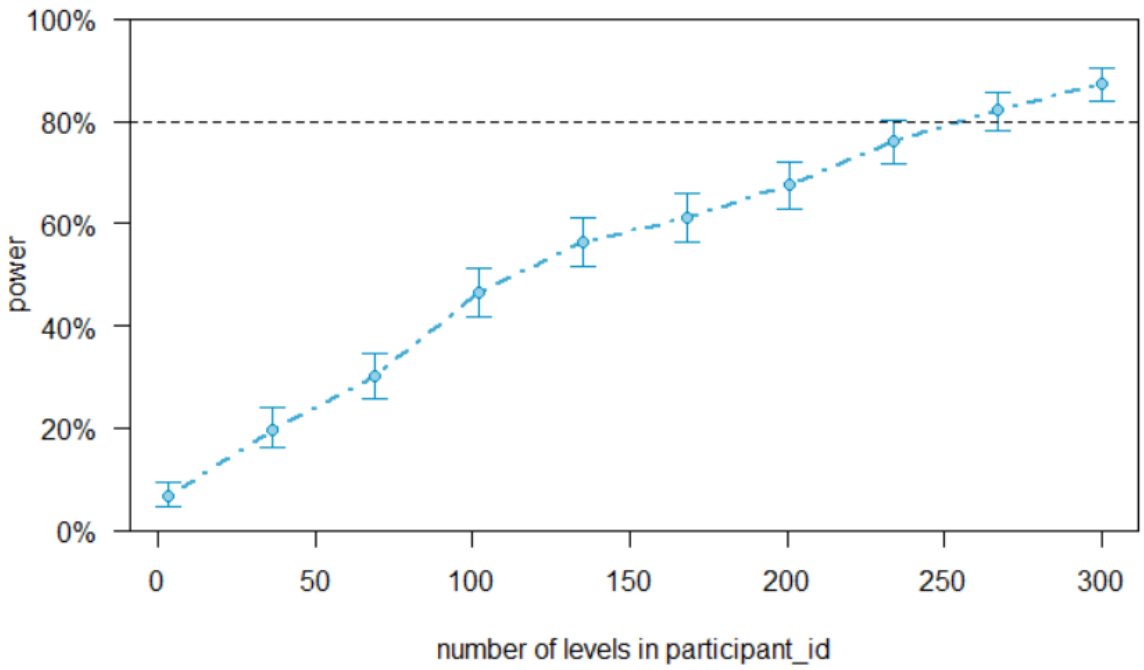
In total, I had 50 news stories with 20 responses each. My first goal was to evaluate the emotionality of the responses to the news stories, with the hope of having every response rated approximately 30 times. I wished to achieve this goal so that in future studies I would be able to choose a subset of existing responses from the ones I had. In each of the 15 trials, I rated on average 7.5 responses (the number of responses varied between 4 and 10). In order to rate all 1000 responses, I needed at least 285 participants. I planned to recruit 300 participants as the selection of the responses was purely random, and I wanted to increase the chance that every response was indeed rated at least 30 times.

My second goal was to estimate the power needed to find the amplification effect. In order to estimate the number of participants required to detect the amplification effect, I conducted a power analysis based on a similar project. In this project, participants rated the average immorality of research institutions based on sequences of experiments conducted by these research institutions. To evaluate the sample size of that project, I conducted a pilot study using 30 trials with 60 participants (https://osf.io/gv92q/). I observed a mean difference between participants' estimations and the actual sequence mean of 0.17 (b = 0.16, SE = .074, t (3019.99) = 2.23, p = .026). Using the simr package, I estimated that the observed power from the previous study was around 53% (42.74% - 63.02%). To achieve more than 80% (83.85% - 90.58%) power, assuming a similar effect size, I would need 300

participants to complete 15 trials (see Figure 5.2S).

However, it is important to note that the previous studies are not a clear comparison to the current design. In those projects, participants did not evaluate each item presented in the sequence before evaluating the overall average. By explicitly instructing participants to estimate each item individually in this project, I might increase the salience of each item and potentially reduce the chance of detecting amplification. Therefore, the power analysis, which was based on a different project, might overestimate the effect size of this project.

Figure 5.2S *Power Analysis for Amplification Effect*



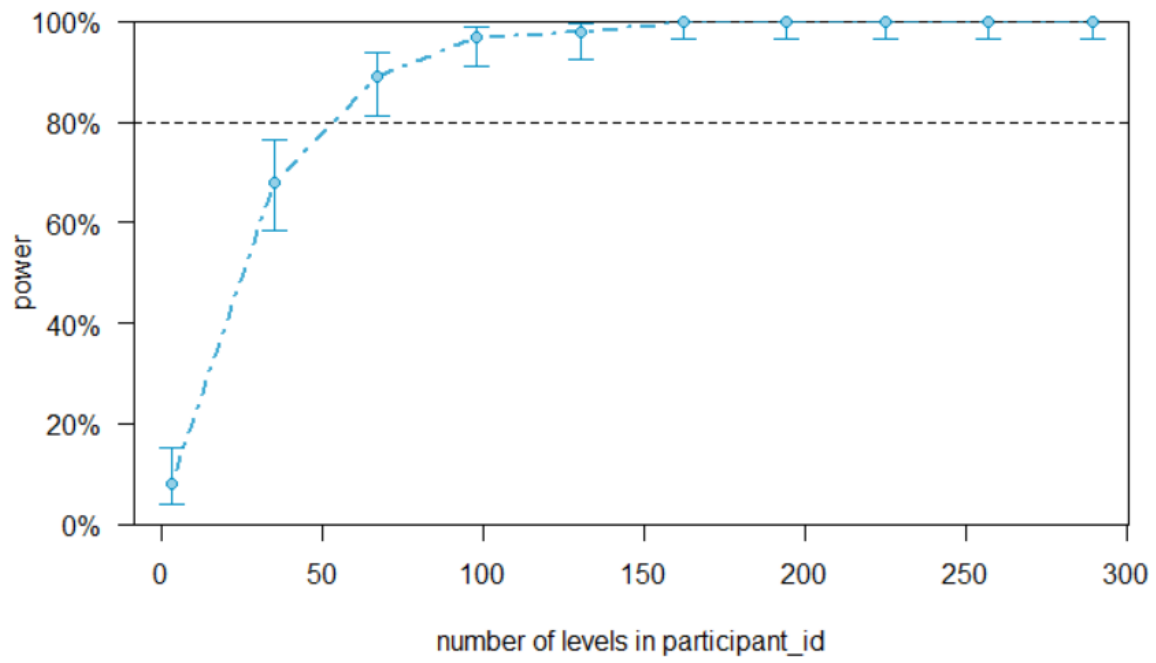
Experiment 2

To evaluate the sample size for the main hypothesis (Q1: Amplification) of this study, I used the results of the previous study conducted in this project

(https://osf.io/s5q3e/). I observed a mean difference between participants' estimations and the actual sequence mean of 1.86 (b = 1.86, SE = .30, t (8324.52) = 6.20, p < .001). Using the simr package, I estimated that the observed power from the previous study was around 98% (96.38% - 100%). To achieve more than 80% power, assuming a similar effect size, I would need 60 participants to complete 15 trials (see Figure 5.3S).

However, I did not know if the amplification effect will be smaller when the news article is presented in the end of the trial instead of the beginning. Due to this uncertainty I aimed for a power near to 100%. I therefore planned to recruit 150 participants to be able to find a slightly smaller effect than in the previous study.

Figure 5.3S *Power Analysis for Amplification Effect*

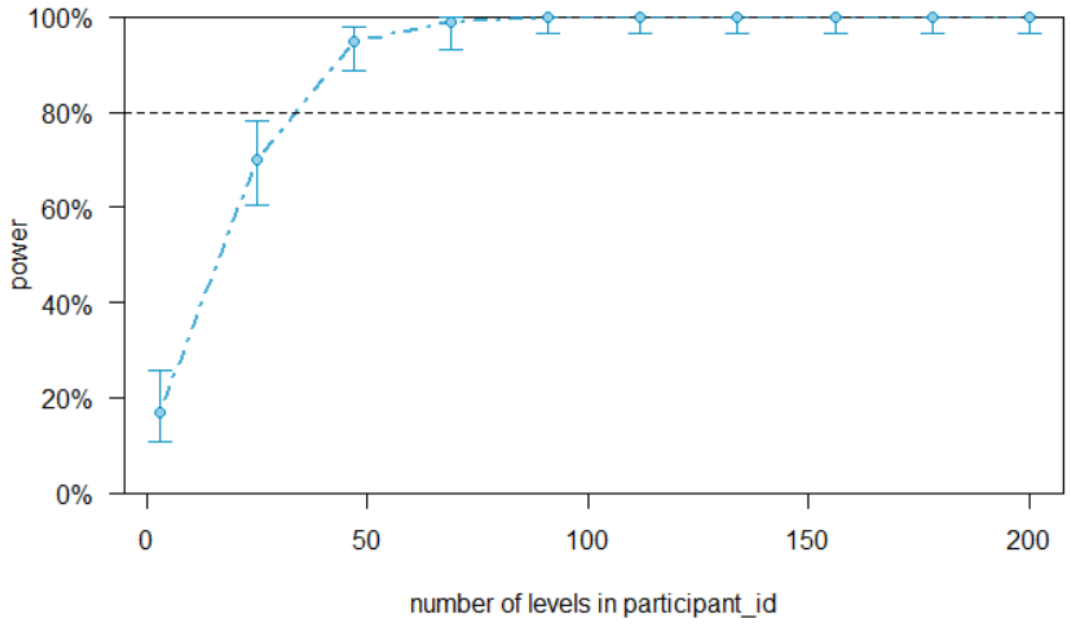


Experiment 3a & b

To evaluate the sample size for the main hypothesis (**Q1: Amplification**) of this study I used the results of the previous study in this project (https://osf.io/s5q3e/). For the calculation I used sequences that showed 8 text responses per sequence to be more comparable to this study. I observed a mean difference between participants' estimations and the actual sequence mean of 2.79 ($b = 2.79$, $SE = 0.84$, $t(462.85) = 3.28$, $p < .001$). Using the *simr* package I estimated that the observed power from the previous study was around 92% (84.84% – 96.48%). In order to achieve more than 80% power assuming a similar effect size I would need 50 participants doing 15 trials (see Figure 5.4S).

However, based on previous projects (https://osf.io/ruc2x/), I expected that the amplification effect will be smaller when the number of stimuli does not vary between sequences. Due to this uncertainty I aimed for a power near to 100%. I therefore planned to recruit 150 participants to be able to find a slightly smaller effect than in the previous study.

Figure 5.4S *Power Analysis for Amplification Effect*



Trial Structure, News Articles and Responses

Each trial consists of three distinct parts, as illustrated in Figure 5.5S (schematic model) and further elaborated in Chapter 5. Initially, participants are required to rate a news story article. Subsequently, they provide a sequence of 4 to 10 individual responses to the aforementioned news article, which are then rated individually (referred to as individual response rating). It is important to note that if the story in the news article is negative all the responses will range from neutral to extremely negative, there will be no positive text in response to a negative text in vise versa (see Figure 5.6S for two examples). Finally, participants rate the average emotional intensity of the responses generated in relation to the news article. For convenience, Table S5.1 presents all negative news articles and their corresponding responses, while Table S5.2 displays all positive articles.

Figure 5.5S *Flowchart of Single Trial*

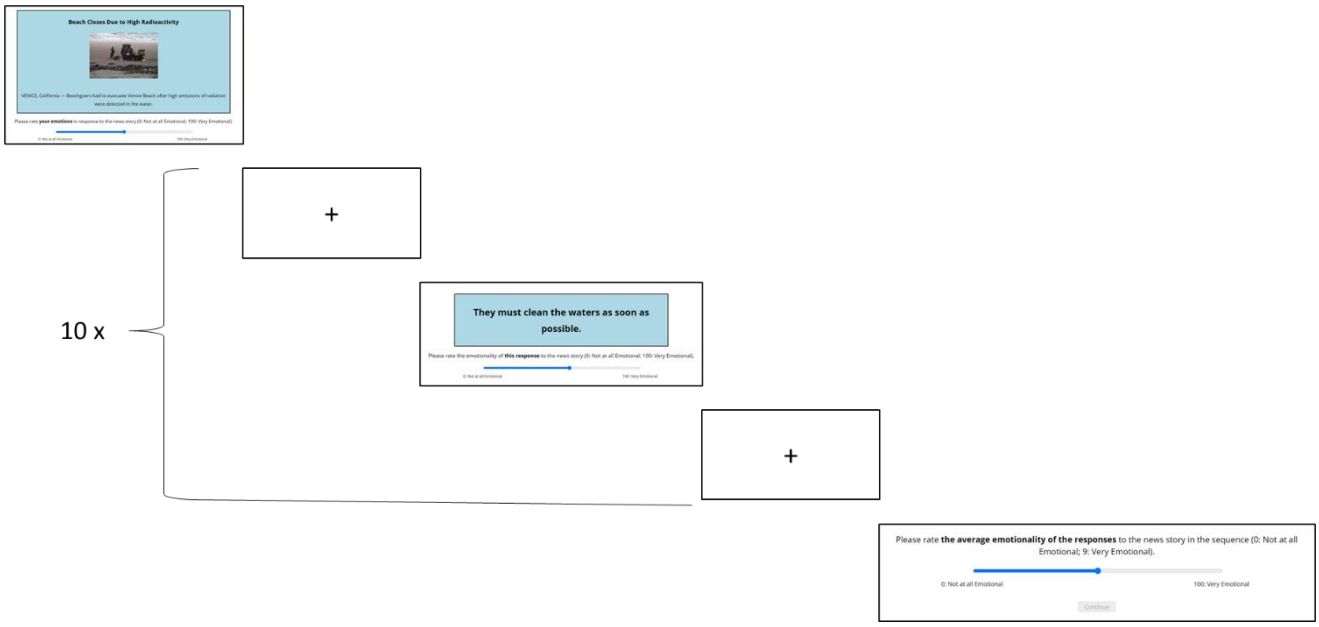


Figure 5.6S *Example of News Article and Response*

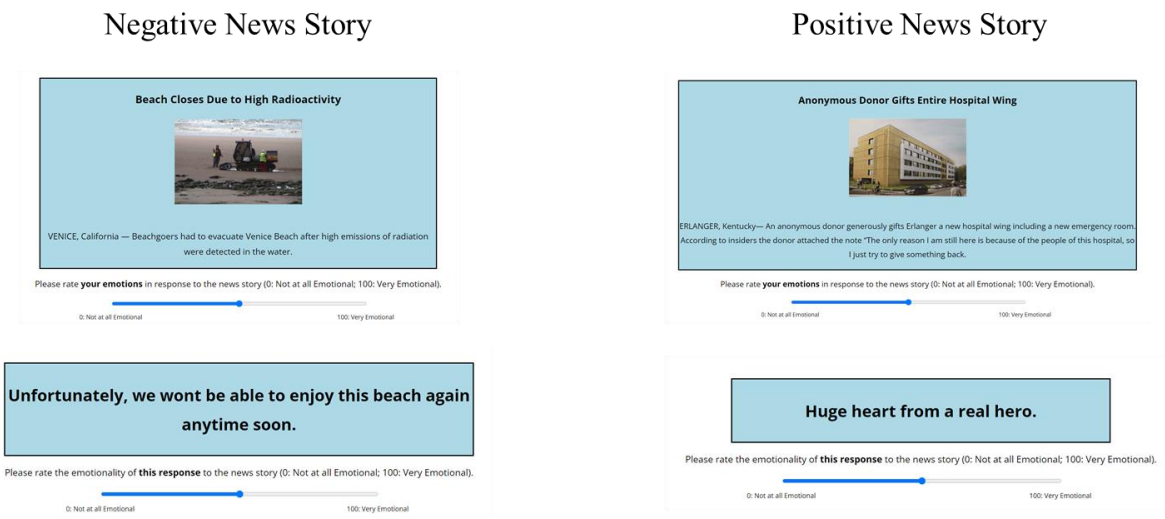


Table 5.1S *Overview and descriptives of negative stories.*

story	valence	M	SD	No. Ratings	Scaled intensity
Bakersfield, Calif. (KGET) — The community paid respects Monday evening to Oildale branch library manager Allison Rodgers, age 27, whose body was discovered in a single car accident that occurred over the weekend.	Neg	68.92	24.59	787	0.20
ST. LOUIS, Missouri — A cable sparked and started the fire that spread through over half the complex, leaving the evacuated tenants helpless watching on as the flames took everything they owned.	Neg	74.14	22.04	708	-0.63
MCALEEN, Texas — The construction of a bridge spanning across the State Highway 107 took an unfortunate turn as officials realized that the bridge cannot be opened for public due to bad concrete quality.han 10m$ of Taxpayers' Money Wasted on Failing Bridge	Neg	48.24	28.88	692	1.60
TUPELO, Mississippi— A rare type of salamander cannot be found anymore around the Tupelo area. Biologists suspect that the local heavy industry contributed to their extinction by disposing toxins into rivers nearby.	Neg	59.49	28.54	590	1.49
SAVANNAH, Georgia— Police warns that a group of scammers try to trick elderly people into sending them money by pretending to be a family member in need. Many victims lost a substantial amount of money this way in last month.	Neg	66.74	27.56	714	1.17
PENSACOLA, Florida— Citizen anger grows as bikes continue to vanish in all locations of the town. Pressure for the police chief increases as they still have no suspects linked to the large number of cases.	Neg	44.77	26.63	701	0.87
BETHESDA, Maryland — Another playground became victim to vandalism last night. A witness states that a group of drunken teenagers might be responsible for this act of violence.	Neg	55.77	26.51	758	0.83

CHICAGO, Illinois — Jared Wyatt, age 33, was arrested and charged with an armed robbery attempt in a family home that occurred last week, officers say.	Neg	49.13	25.85	717	0.61
AUSTIN, Texas — Police documents reveal that the motorist who ran over and killed a bicyclist last month did so in a stolen car.	Neg	75.16	20.03	717	-1.29
WICHITA, Kansas — According to witnesses shots were fired in the ongoing hostage situation in a local restaurant. The police were thus far not successful in their negotiations with the perpetrator.	Neg	58.99	25.39	704	0.46
NEW YORK, New York — Experts determined what was believed to be an Albrecht Dürer artwork that sold for $1.2 million dollars at an auction last weekend was in fact forged.	Neg	29.32	24.87	679	0.29
LOUISVILLE, Kentucky — A worker for Cummings Steel died due to injuries sustained during an accident while working on construction for Planz Elementary, according to representatives from Cummings Steel.	Neg	69.63	20.47	808	-1.14
CONCORD, New Hampshire — Charles Brinkley, age 27, died in a shoot-out late last night in front of Haines library, after responding to a call about an armed robbery in the area.	Neg	77.67	19.60	635	-1.43
ASPEN, Colorado — The body of avid hiker Greg Sumpter, age 47, was found buried under snow on a trail in Aspen late Tuesday night after the man went missing over the weekend.	Neg	70.02	22.44	793	-0.50
BAKERSFIELD, California — Students and staff at South High School were shocked and saddened to discover that music teacher Adam Jameson, who taught at the school for forty-two years, died of a heart attack late Sunday night.	Neg	73.34	19.06	639	-1.60
VENICE, California — Beachgoers had to evacuate Venice Beach after high emissions of radiation were detected in the water.	Neg	56.87	27.55	776	1.17
RENO, Nevada — Hawk numbers have declined drastically, researchers say, due to illegal hunting without license in the area.	Neg	61.72	26.30	695	0.76

HARRISBURG, Pennsylvania — Lena Smith, 28, was arrested Wednesday for running a scam charity for the homeless, police reports say.	Neg	64.23	22.67	679	-0.43
WASHINGTON, D.C. — Intelligence officials claim a new cryptocurrency was mainly used to launder money by those in the illegal drug trade.	Neg	36.14	24.40	799	0.14
AURORA, Colorado — Calls for Mayor Hall's resignation have grown louder in the wake of allegations he diverted taxpayer dollars to the businesses of personal friends.	Neg	51.68	26.72	769	0.90
NORFOLK, Virginia — Homeowners displaced at the city's request so 24th street could be widened have filed a lawsuit claiming they haven't been properly reimbursed for moving, leaving many unable to remain in new homes.	Neg	62.21	25.63	744	0.54
TALLAHASSEE, Florida — A man who shot his neighbor during a confrontation that took place on the man's front yard was acquitted of murder under the state's Stand Your Ground laws.	Neg	69.50	22.27	750	-0.56
ANN ARBOR, Michigan — A local high school football coach was charged with supplying his athletes performance-enhancing drugs throughout the season.	Neg	57.80	26.09	722	0.69
BOISE, Idaho — A shooting at a local mall leaving a victim in critical condition was due to personal animosity and was not an act of terrorism, according to police reports.	Neg	71.39	22.44	654	-0.50
TAFT, California — Taft's homeless shelter will be forced to permanently close on Monday, potentially leaving many without food and refuge from the heat as summer looms.	Neg	75.07	24.29	736	0.10

Table 5.2S. *Overview and descriptives of positive stories.*

story	valence	M	SD	No. Ratings

FOND DU LAC, Wisc. - Mario Chalmers, age 32, fulfilled the promise he made to himself while undergoing chemotherapy by completing the Boston Marathon.	Pos	72.84	23.47	717
SYRACUSE, NY - A weekend-long search reached a happy conclusion when missing child Ethan Aguirre, age 7, was found in the woods near his home and returned to his family.	Pos	79.94	17.74	712
DES MOINES, Iowa - Private First Class in the National Guard Evan Khiev, age 31, overcame two flight delays home from deployment in Afghanistan to arrive just in time to witness the birth of daughter Amara.	Pos	72.90	23.19	717
VISALIA, Calif. - Kevin Delgado of Delgado Engineering, age 43, donated enough money to house an extra hundred animals in the Tulare County Animal Shelter for the year.	Pos	74.68	21.62	788
NEW YORK, New York - A group of subway passengers saved a man who fell onto the tracks mere moments before their train arrived, no doubt saving his life in the process.	Pos	78.30	16.96	770
LAUREL, Delaware- Kellogg Highschool was able to secure a historical place in the final for the National Science Championship where they will continue to make their families and community proud.	Pos	44.19	27.14	641
WASILA, Alaska- James Carson, Wasilla firefighter, heroically rescued a deer that broke into a frozen river by barehandedly pulling it out of its trap.	Pos	71.46	25.22	820
ERLANGER, Kentucky- An anonymous donor generously gifts Erlanger a new hospital wing including a new emergency room. According to insiders the donor attached the note -The only reason I am still here is because of the people of this hospital, so I just try to give something back.	Pos	77.77	19.30	770
HOBOKEN, New Jersey- Archaeologists discovered the remains of a boat that might date back to the time of the Vikings. If this proves to be true, this could be some of the earliest evidence of European settlers on the east coast of the USA.	Pos	40.15	28.43	671

BATON ROUGE, Louisiana - High school junior Elton Pierce delivered a memorable performance, leading all scorers and scoring the game-winner in the boys. basketball state championship game a day after losing his father to pancreatic cancer.	Pos	76.07	20.95	755
WEST YELLOWSTONE, Montana - Charlisa Hentley, age 33, rescued her two-year-old daughter from drowning in their backyard swimming pool.	Pos	74.63	21.66	672
SANTA FE, New Mexico - 40 year old Rosa Marques, who was blinded due to a rare genetic disorder, successfully underwent an experimental surgery to regain her vision.	Pos	75.61	20.88	768
TRENTON, New Jersey - Martial Artist Jane Olstead, age 27, put her Jiu-Jitsu training to use Thursday evening when she disarmed and detained a robber wielding a knife in a liquor store, keeping the perpetrator pinned down until police arrived.	Pos	59.72	23.54	730
SPRINGFIELD, Illinois - Local businessman Alfred Shipes stayed true to his word, donating $1,000 for every touchdown scored in his son's high school football league. The team scored 40 in total this season.	Pos	61.25	22.32	732
ALBUQUERQUE, New Mexico - Officers found and arrested the man who kidnapped an eight-year-old boy two hours after the kidnapping took place, returning the boy unharmed to his parents.	Pos	76.45	20.39	763
LOS ANGELES, California -The CEO of the post-production company Drifters, Andrew Frausto, age 37, donated his suit collection to men at the homeless shelter to help them interview for jobs.	Pos	68.45	25.54	701
SYRACUSE, New York - The Orange High School choir sang to seven different nursing homes yesterday as part of their annual tradition, where they invite senior citizens to sing along with them.	Pos	53.55	28.28	636
MINNEAPOLIS, Minnesota - Local non-profit Walk Don't Drive succeeded in cutting emissions and also cut down fatal traffic accidents by encouraging commuters to walk to work, studies show.	Pos	46.65	24.83	756

ARLINGTON, Texas - A man claims his dog woke him from his nap in time for him to evacuate safely, call the fire department, and have them save his house from major damage.	Pos	73.18	23.57	657
BOSTON, Massachusetts - Jennifer Danes, 27, will star in tonight's opening of Swan Lake after a yearlong rehabbing of a leg injury.	Pos	41.95	25.48	807
NEW HAVEN, Connecticut - Brandon Wakefield, 32, overcame a heroin addiction and worked her way through community college then through UCLA before earning her Ph.D. from Yale this semester.	Pos	59.43	25.39	605
PORTLAND, Oregon - High school seniors Jacob and James O'Connor were both drafted by the Seattle Mariners yesterday in the 8th and 9th picks of the Major League Baseball draft.	Pos	36.65	28.22	729
BOSTON, Massachusetts - The bee population in North America is increasing according to a study from Harvard University. This positive trend sparks hope for the future of this plane.	Pos	54.99	28.12	813

Amplification and Potential Moderators (Correlation with Scales)

I examined if the amplification effect was moderated by demographic variables such as their gender and race or their individual differences as assessed by a few scales: social interaction anxiety scale (SIAS-6, Mattick & Clarke, 1998), PERS reactivity scale (Becerra, Preece, Campitelli, & Scott-Pillow, 2019), and the Big-Five personality scale (Gosling, Rentfrow, & Swann Jr, 2003). Throughout my analysis, I calculated a difference score by comparing my estimation of the objective mean of the sequence, as determined by participants individual response ratings, to their sequence rating. Subsequently, I performed a series of mixed model analyses, utilizing each potential moderator as a predictor and the difference score as the outcome variable. Consistent with my previous analysis, I incorporated by-story and by-participant random variables.

Demographics as Moderators for Amplification

Race. Participants were asked to indicate their race from 5 options: Black, Hispanic, White, Asian and other, which they could specific individually. Participants could select multiple options for this question. As most participants were white, I binarized the options into "white" and "other races" participants to increase the possibility of finding a moderator effect. The results displayed in Table 5.4S shows the coefficient for being white compared to identifying oneself as any of the other races.

Table 5.3S *Overview of number of participants identifying with a certain race.*

Study	***N***	**Black**	**Hispanic**	**White**	**Asian**	**Other**
1: Article than response	282	57	44	185	6	17
2: Responses than article	139	18	37	77	2	5
3a: Memory	138	5	3	107	10	13
3b: Representativeness	146	11	2	114	8	11

Table 5.4S *The coefficients of race (binarized race as "white" and "other races") predicting participants tendency to amplify (difference between actual intensity of sequence and participants estimation).*

Study	**b [ci], (se)**	**t (df)**	**p**
1: Article than response	-0.37 [-1.46 – 0.71], (0.55)	-0.68 (4224)	.50
2: Responses than article	-1.07 [-2.44 – 0.31], (0.70)	-1.52 (2069)	.13
3a: Memory	3.35 [0.84 – 5.85], (1.28)	2.62 (2056)	.009**
3b: Representativeness	0.13 [-2.32 – 2.58], (1.25)	0.10 (2179)	.91

Notes. *The coefficient shows the difference of being "white" compared to identifying themselves as any "other race".*

Gender. Participants were asked to indicate their gender from 4 options: male, female, other or prefer not to say. The results displayed in Table 5.6S shows the coefficient for being female compared to identifying oneself as male or other.

Table 5.5S. *Overview of number of participants identifying with a certain race.*

Study	***N***	**Male**	**Female**	**Prefer not to say**	**Other**
1: Article than response	282	145	135	1	1
2: Responses than article	139	67	70	1	1
3a: Memory	139	70	68	-	-
3b: Representativeness	146	71	75	-	-

Table 5.6S. *Model coefficients of gender predicting participants tendency to amplify (difference between actual intensity of sequence and participants estimation).*

Study	**b [ci], (se)**	**t (df)**	**p**
1: Article than response	0.84 [-0.23 – 1.91], (0.55)	1.53 (4224)	.12
2: Responses than article	-1.45 [-2.80 – -0.10], (0.69)	-2.10 (2069)	.04*
3a: Memory	0.07 [-2.07 – 2.22], (1.09)	0.07 (2056)	.94
3b: Representativeness	0.57 [-1.46 – 2.59], (1.03)	0.55 (2179)	.58

Notes. *The coefficient shows the difference of being "female" compared to identifying themselves as "male".*

Individual Differences Scales as Predictors for Amplification

Social Anxiety. I measured social anxiety using an abbreviated version of the social interaction anxiety scale (SIAS-6, Mattick & Clarke, 1998)) which includes the following items on a scale from (0 - Not at all to 4-Extremely)

1. *I have difficulty making eye-contact with others.*
2. *I find it difficult mixing comfortably with the people I work with.*
3. *I tense-up if I meet an acquaintance on the street.*
4. *I feel tense if I am alone with just one person.*
5. *I have difficulty talking with other people.*
6. *I find it difficult to disagree with another's point of view.*

Table 5.7S. *Ns, means, and standard deviations of the Social Anxiety Questionnaire (SIAS-6).*

Study	***n***	***M***	***SD***
1: Article than response	282	1.70	0.63
2: Responses than article	139	1.86	0.62
3a: Memory	138	1.67	0.73
3b: Representativeness	146	1.72	0.78

I then used multilevel models to predict amplification (difference between estimation and actual mean of the sequence) by their social anxiety score. I again added a random intercept of participants' id to the model.

Table 5.8S. *Model coefficients of social anxiety score predicting tendency to amplify (difference between actual intensity of sequence and participants estimation).*

Study	**b [ci], (se)**	**t (df)**	**p**
1: Article than response	-0.45 [-1.76 – 0.02], (0.45)	-1.92 (3834)	.055
2: Responses than article	0.53 [-0.57 – 1.62], (0.56)	0.94 (2069)	.34

3a: Memory	1.38 [-0.07 – 2.82], (0.74)	1.87 (2056)	.062
3b: Representativeness	0.95 [-0.34 – 2.23], (0.65)	1.45 (2179)	.15

Notes. *The coefficient shows the correlation coefficient of social anxiety score.*

Personality. I assessed participants personality using the ten item personality scale (TIPI, Gosling et al., 2003). Each of the five personality factors has two items in this scale. One of the items is always reverse coded. Participants rate themselves on how much these two items apply to themselves on a scale from 1- Strongly Disagree to 7-Strongly Agree. The score is the average of two items (after reversing one item).

Extraversion: Is being assessed by the items:

1. *Extraverted, enthusiastic*
2. *Reserved, quiet*

Table 5.9S. *Ns, means, and standard deviations of the Big Five Personality Questionnaire Factor Extraversion.*

Study	***n***	***M***	***SD***
1: Article than response	282	3.73	1.65
2: Responses than article	139	3.55	1.51
3a: Memory	138	3.38	1.56
3b: Representativeness	146	3.34	1.56

I then used multilevel models to predict amplification (difference between estimation and actual mean of the sequence) by their extraversion score. I again added a random intercept for participants.

Table 5.10S. *The coefficients of the moderator (Big Five extraversion score) and participants tendency to amplify (difference between actual intensity of sequence and participants estimation) .*

Study	**b [ci], (se)**	**t (df)**	**p**

1: Article than response	0.10 [-0.23 - 0.42], (0.17)	0.60 (4224)	.55
2: Responses than article	-0.14 [-0.59 – 0.32], (0.23)	-0.60 (2069)	.55
3a: Memory	-0.66 [-1.34 – 0.02], (0.35)	-1.91 (2056)	.056
3b: Representativeness	-0.11 [-0.76 – 0.54], (0.33)	-0.35 (2179)	.73

Notes. *The coefficient shows the correlation coefficient of extraversion score.*

Neuroticism: Is being assessed by the items:

1. *Anxious, easily upset.*
2. *Calm, emotionally stable.*

Table 5.11S. *Ns, means, and standard deviations of the Big Five personality questionnaire factor neuroticism.*

Study	***n***	***M***	***SD***
1: Article than response	282	3.64	1.45
2: Responses than article	139	4.03	1.38
3a: Memory	138	3.63	1.60
3b: Representativeness	146	3.63	1.60

I then used multilevel models to predict amplification (difference between estimation and actual mean of the sequence) by their neuroticism score. I again added a random intercept for participants' id.

Table 5.12S. *The coefficients of the moderator (Big Five neuroticism score) and participants tendency to amplify (difference between actual intensity of sequence and participants estimation).*

Study	**b [ci], (se)**	**t (df)**	**p**

1: Article than response	-0.44 [-0.80 – -0.07], (0.19)	-2.35 (4224)	.019*
2: Responses than article	0.19 [-0.31 – 0.68], (0.25)	0.73 (2069)	.46
3a: Memory	0.12 [-0.55 – 0.79], (0.34)	0.36 (2056)	.72
3b: Representativeness	0.54 [-0.09 – 1.16], (0.32)	1.68 (2179)	.094

Notes. *The coefficient shows the correlation coefficient of neuroticism score.*

Agreeableness: Is being assessed by the items:

1. *Critical, quarrelsome.*
2. *Sympathetic, warm.*

Table 5.13S. *Ns, means, and standard deviations of the Big Five Personality questionnaire factor agreeableness.*

Study	***n***	***M***	***SD***
1: Article than response	282	3.27	1.31
2: Responses than article	139	4.03	1.38
3a: Memory	138	2.95	1.21
3b: Representativeness	146	2.97	1.18

I then used multilevel models to predict amplification (difference between estimation and actual mean of the sequence) by their agreeableness score. I again added a random intercept for participants' id.

Table 5.14S. *The coefficients of the moderator (Big Five agreeableness score) and participants tendency to amplify (difference between actual intensity of sequence and participants estimation).*

Study	**b [ci], (se)**	**t (df)**	**p**

1: Article than response	0.16 [-0.31 – 0.63], (0.24)	0.66 (4224)	.50
2: Responses than article	0.07 [-0.57 – 0.72], (0.33)	0.22 (2069)	.44
3a: Memory	0.36 [-0.53 – 1.25], (0.45)	0.80 (2056)	.42
3b: Representativeness	-0.10 [-0.96 – 0.76], (0.44)	-0.24 (2179)	.81

Notes. *The coefficient shows the correlation coefficient of agreeableness score.*

Conscientiousness: Is being assessed by the items:

1. *Dependable, self-disciplined.*
2. *Disorganized, careless*

Table 5.15S. *Ns, means, and standard deviations of the Big Five personality questionnaire factor Conscientiousness.*

Study	***n***	***M***	***SD***
1: Article than response	282	4.90	1.42
2: Responses than article	139	4.68	1.24
3a: Memory	138	5.11	1.23
3b: Representativeness	146	5.10	1.38

I then used multilevel models to predict amplification (difference between estimation and actual mean of the sequence) by their conscientiousness score. I again added a random intercept for participants.

Table 5.16S. The coefficients of the moderator (Big Five conscientiousness score) and participants tendency to amplify (difference between actual intensity of sequence and participants estimation).

Study	**b [ci], (se)**	**t (df)**	**p**

1: Article than response	0.25 [-0.13 – 0.62], (0.19)	1.30 (4224)	.19
2: Responses than article	0.03 [-0.52 – 0.58], (0.28)	0.10 (2069)	.92
3a: Memory	-0.69 [-1.56 – 0.17], (0.44)	-1.57 (2056)	.11
3b: Representativeness	-0.49 [-1.22 – 0.23], (0.37)	-1.33 (2179)	.18

Notes. *The coefficient shows the correlation coefficient of conscientiousness score.*

Openness: Is being assessed by the items:

I. *Open to new experiences, complex.*
II. *Conventional, uncreative.*

Table 5.17S. *Ns, means, and standard deviations of the Big Five personality questionnaire factor openness.*

Study	***n***	***M***	***SD***
1: Article than response	282	5.10	1.27
: Responses than article	139	4.89	1.19
3a: Memory	138	4.67	1.26
3b: Representativeness	146	4.85	1.18

I then used multilevel models to predict amplification (difference between estimation and actual mean of the sequence) by their openness score. I again added a random intercept for participants.

Table 5.18S. *The coefficients of the moderator (Big Five openness score) and participants tendency to amplify (difference between actual intensity of sequence and participants estimation).*

Study	**b [ci], (se)**	**t (df)**	**p**

1: Article than response	0.16 [-0.26 – 0.58], (0.21)	1.10 (4224)	.49
2: Responses than article	-0.05 [-0.62 – 0.53], (0.29)	-0.16 (2069)	.87
3a: Memory	-0.06 [-0.91 – 0.79], (0.43)	-0.13 (2056)	.89
3b: Representativeness	-0.20 [-1.06 – 0.66], (0.44)	-0.46 (89.61)	.64

Notes. *The coefficient shows the correlation coefficient of openness score.*

Positive and Negative Emotional Reactivity. The Perth Emotional Reactivity Scale (PERS) is used to assess individuals' emotional reactivity, because their reactivity might influence how much people might overestimate the emotionality of a sequence. The scale consists of a series of statements or items to which participants respond based on their own experiences. Participants rate the degree to which they agree or disagree with each statement, indicating the intensity of their emotional reactions. The PERS comprises a total of 18 items, with 9 items measuring positive emotional reactivity and 9 items measuring negative emotional reactivity. Participants rate their current level of emotional experience on a scale ranging from 1. Very slightly/not at all to 5. Extremely.

Positive Emotional Reactivity: The 10 items assessing positive emotional reactivity on the PERS are as follows:

1. *I tend to get happy very easily*
2. *When I'm happy, the feeling stays with me for quite a while*
3. *When I am joyful, I tend to feel it very deeply*
4. *I feel good about positive things in an instant*
5. *When I'm feeling positive, I can stay like that for a good part of the day*
6. *I experience positive mood very strongly*
7. *I react to good news very quickly*
8. *I can remain enthusiastic for quite a while*
9. *When I'm enthusiastic about something, I feel it very powerfully*

Table 5.19S. *Ns, means, and standard deviations of the Positive Emotional Reactivity Category from the PERS.*

Study	***n***	***M***	***SD***
1: Article than response	282	3.61	0.84
2: Responses than article	139	3.51	0.76
3a: Memory	138	3.41	0.90
3b: Representativeness	146	3.42	0.89

I then used multilevel models to predict amplification (difference between estimation and actual mean of the sequence) by their positive reactivity score. I again added a random intercept for participants.

Table 5.20S. The coefficients of the moderator (PERS positive reactivity score) and participants tendency to amplify (difference between actual intensity of sequence and participants estimation).

Study	**b [ci], (se)**	**t (df)**	**p**
1: Article than response	0.12 [-0.51 – 0.75], (0.32)	0.38 (4224)	.70
2: Responses than article	0.10 [-0.80 – 0.99], (0.46)	0.21 (2069)	.83
3a: Memory	0.17 [-1.01 – 1.36], (0.60)	0.28 (2056)	.77
3b: Representativeness	0.52 [-0.61 – 1.64], (0.57)	0.90 (2179)	.37

Notes. *The coefficient shows the correlation coefficient of positive affect score.*

Negative Emotional Reactivity: The 9 items assessing positive emotional reactivity on the PERS are as follows:

1. *I tend to get upset very easily*
2. *When I'm upset, it takes me quite a while to snap out of it*
3. *If I'm upset, I feel it more intensely than everyone else*
4. *I tend to get disappointed very easily*
5. *It's hard for me to recover from frustration*
6. *Normally, when I'm unhappy I feel it very strongly*
7. *I tend to get pessimistic about negative things very quickly*
8. *Once in a negative mood, it's hard to snap out of it*
9. *My negative feelings feel very intense*

Table 5.21S. *Ns, means, and standard deviations of the Negative Emotional Reactivity Category from the PERS.*

Study	***n***	***M***	***SD***
1: Article than response	282	3.61	0.84
2: Responses than article	139	3.69	0.78
3a: Memory	138	3.54	0.98
3b: Representativeness	146	3.66	1.03

I then used multilevel models to predict amplification (difference between estimation and actual mean of the sequence) by their negative affect score. I again added a random intercept for participants.

Table 5.22S. *The coefficients of the moderator (PERS negative reactivity score) and participants tendency to amplify (difference between actual intensity of sequence and participants estimation).*

Study	**b [ci], (se)**	**t (df)**	**p**
1: Article than response	-0.53 [-1.24 – 0.19], (0.36)	-1.44 (4224)	.15
2: Responses than article	0.26 [-0.62 – 1.13], (0.45)	0.57 (2069)	.57
3a: Memory	-1.05 [-2.12 – 0.02], (0.55)	-1.92 (2056)	.055

3b: Representativeness	0.22 [-0.75 – 1.20], (0.50)	0.45 (2179)	.65

Notes. *The coefficient shows the correlation coefficient of negative affect score.*