ABSTRACT

HAQUE, AMANUL. News, Networks, and Narratives: A Language Model and Social Simulation Approach. (Under the direction of Dr. Munindar P. Singh).

News has a profound influence on public opinion, shaping both short-term discourse and long-term social attitudes. Social media has amplified this influence and has made public journalism more accessible than ever, creating a news overload. Moreover, the rise of social media as a news source has raised concerns about polarization and echo chambers.

In this dissertation, we present computational frameworks to analyze news and social media discourse based on news.

First, we analyze political slant in election-related news and discourse on social media based on sentiment toward political figures. We observe that both traditional online news and social media exhibit political slant, with significant differences across publishers with different presumed political leanings. However, content on social media is more sentiment-driven, with a stronger political slant than traditional news. User responses on social media reveal a moral divide between readers of different news sources, suggesting that audiences focus on different aspects of the same news depending on which news source they follow.

Second, we identify the moral framing in the news related to the ongoing Gaza conflict and analyze it using the well-known war and peace journalism framework. We employ a vector subspace projection approach to identify moral framing and leverage language models fine-tuned on a crowdsourced annotated dataset for war and peace journalism analysis. We observe significant variations in how different publishers frame the same events, with distinct moral perspectives aligning with their presumed political leanings, suggesting partisan coverage. A concerning trend we observe is the causal association between news publishers, suggestive of agenda-setting, where one publisher's framing influences another's.

Third, we investigate the effects of selective exposure to congenial (like-minded) content on polarization dynamics in a social network via a study on a custom-built simulation testbed. We observe that whereas selective exposure to congenial content enhances user satisfaction by tailoring content to individual preferences, it contributes to ideological divides, intensifying polarization within the network. Moreover, it limits the reach of content across diverse audiences. In contrast, increased user tolerance (for noncongenial content) mitigates polarization; however, this comes at the expense of user satisfaction.

Our analysis reveals slants in political news and cross-publisher influence on moral framing, combined with a moral divide on social media across audiences of different news sources, raising concerns about agenda setting and polarization. Furthermore, our simulation study

demonstrates that the risk of polarization is greatest when selective exposure is high and user tolerance is low.

News, Networks, and Narratives: A Language Model and Social Simulation Approach

by
Amanul Haque

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Computer Science

Raleigh, North Carolina
2025

APPROVED BY:

Dr. Collin F. Lynch	Dr. Christopher Healey
Dr. Andrew Davis	Dr. Munindar P. Singh Chair of Advisory Committee

BIOGRAPHY

Amanul was born in Varanasi (Benaras), which is considered one of the world's oldest continuously inhabited cities. Mark Twain aptly described it, *"Benaras is older than history, older than tradition, older even than legend, and looks twice as old as all of them put together."* Amanul spent his formative years in Varanasi and completed his schooling there.

In 2015, he earned a Bachelor of Engineering in Information Science from PES University, Bangalore, and joined Oracle as a Member of Technical Staff.

In 2017, he moved to the United States to pursue a Master of Science in Computer Science at North Carolina State University. Under the advisement of Dr. Collin F. Lynch, he completed his thesis and graduated in 2019. He then continued at NC State University for a Ph.D. in Computer Science, working under Dr. Munindar P. Singh.

During his Ph.D., Amanul was recognized for his contributions to teaching and leadership by NC State University. He received the "Outstanding TA Award" (2021) for his contributions to "Natural Language Processing" and "Social Computing & Decentralized AI" courses. In 2023, he was honored with the "Outstanding Student Leader Award". He served as President and Head of Events for Maitri, the Indian Graduate Student Association at NC State University, from Spring 2022 to Fall 2023.

His industry experience includes internships at Lenovo, Seagate, and Coupang, as well as a research scholar position at the Laboratory of Analytic Science (LAS) during SCADS 2024 (Summer Conference on Applied Data Science).

Beyond academia, Amanul is passionate about Indian classical music and is a self-taught flutist. He enjoys reading non-fiction, particularly biographies and books on human behavior and psychology, and maintains an active lifestyle through workouts and meditation. A nature enthusiast, he loves hiking and exploring the outdoors.

ACKNOWLEDGEMENTS

I am deeply grateful to my advisor, Dr. Munindar P. Singh, for his unwavering support and invaluable guidance throughout my Ph.D. His mentorship has shaped my understanding of research, writing, and attention to detail. He has always encouraged my ideas and inspired me to set ambitious goals. This dissertation would not have been possible without his supervision.

I would also like to sincerely thank my committee members, Dr. Collin F. Lynch, Dr. Christopher Healey, and Dr. Andrew Davis, for their insightful feedback and constructive discussions, which greatly improved this work.

A special thanks to Dr. Nirav Ajmeri, who mentored me in the early stages of my Ph.D. His guidance helped me navigate the fundamentals of research, and his contributions, particularly to my work on multiagent social simulation (Chapter 5), were invaluable. His constant support and direction have been instrumental in my academic journey.

I am also grateful to Dr. Davoud Mougouei, my collaborator on several of my early projects related to news discourse analysis. Our discussions and his mentorship have significantly influenced my research. His belief in me has been a source of confidence and motivation.

I extend my appreciation to my labmates for their support and stimulating discussions. In particular, Dr. Hui Guo, who played a key role in shaping my early research, and Mansi Saxena, whose assistance in crowdsourcing annotations for the war and peace journalism dataset (Chapter 4) was invaluable.

Finally, and most importantly, I am profoundly grateful to my family for their unwavering support. I especially thank my father, Mr. Abdul Haque, who has been my greatest mentor and source of inspiration. His belief in me has shaped who I am today.

This research was partially supported by the National Science Foundation under grant IIS-1908374 and IIS-2116751, and a gift from Facebook.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER

1

NEWS, SOCIAL MEDIA, AND POLITICAL POLARIZATION

1.1 Background and Motivation

1.1.1 News Influence

News media plays a crucial role in shaping public opinion. News informs people about the happenings around the world and influences public discourse on critical social and political issues. News can influence the course of wars (Hallin 1989; Kull et al. 2003; Aday et al. 2005), electoral debates and voter perceptions (Iyengar and Kinder 1989; DellaVigna and Kaplan 2007), policy priorities (McCombs and Shaw 1972), and even public perception of health hazards of products such as cigarettes (Amos and Haglund 2000).

While the topics covered in the news set the agenda for public discourse, the framing of information influences how it is understood. The presentation of news shapes the viewpoint through which the public understands a topic (Entman 1993; Zaller 1992a). People's preferences are influenced substantially by how news is presented and framed (Shah et al. 2002). Contentious social and political issues, such as elections, immigration, and racial injustice, can be framed in ways that either mobilize or polarize public opinions.

The quality of online news plays a critical role in shaping public discourse. Factual and balanced reporting fosters informed discussions and nuanced perspectives, whereas sensationalized news laden with emotionally charged language distorts public opinion by prioritizing shock value over truth. Deviation from journalistic objectivity results in media bias, which can increase political polarization (Gerber et al. 2009).

Although high journalistic standards are desirable, news publishers don't always deliver. News publishers have financial incentives and constraints that influence their reporting (Dunaway 2013). Moreover, they often align their content to the preferences of their readers (Gentzkow and Shapiro 2010; Branton and Dunaway 2009), advertisers (Hamilton 2003), and their own ideology (Hassell et al. 2022; Dunaway 2008), which affects how audiences of different news publishers may perceive the same news event. Additionally, news publishers often resort to sensationalizing news (or presentation) to exploit our preference for negative stimuli. Negative emotions are processed more thoroughly than positive emotions (Baumeister et al. 2001) and have a stronger influence (Ito et al. 1998). Moreover, negativity information spreads faster than positive information (Rozin and Royzman 2001).

1.1.2 Increasing Polarization and Declining Trust

News coverage in recent times has become increasingly polarized, and a larger fraction of mentions of political figures in the news are associated with polarized language (Chinn et al. 2020). Most Americans believe the news is biased (Jones 2018), and they do not trust news media (Survey 2018). A study conducted by Media Insight Project (2018) identified that a large portion of Americans (42%) feel news media looks largely like opinion and commentary. As a consequence of partisan news, polarization in the US is on the rise (Boxell et al. 2020; Dimock and Wike 2021), and the gap between the political values of Democrats and Republicans is now wider than at any point in recent history (Heltzel and Laurin 2020; Pew 2017).

An important contributor to this increased polarization is the social media. Social media has transformed how news is consumed and how public opinion is shaped. It offers immediate access to diverse sources and viewpoints and enables the dissemination of news across global audiences almost instantly. Moreover, people are moving towards using social media platforms to keep up with the news. For many people, especially young adults, social media is one of the most common sources of news, particularly political news (Infield 2020). The popularity of social media for political discussions has led many politicians to use social media platforms to set the agenda for discussion that favors their political interests (Yang et al. 2016). Moreover, studies suggest that adults who use social media sites as a political tool are likelier to participate in politics offline (Bode et al. 2014).

On the one hand, the enhanced connectivity has empowered readers with quick access to real-time news; on the other hand, it presents challenges in discerning credible, high-quality reporting from misinformation and sensationalized content. The effects of negativity can be seen in online news consumption (Robertson et al. 2023). Emotionally charged negative language in online content correlates with higher user engagement and sharing (Pröllochs et al. 2021b,a; Brady et al. 2017; Berger and Milkman 2012; Chuai and Zhao 2022).

The ability to connect with like-minded users and filter content on social media deepens the divide in news exposure across different audiences. Individuals are more likely to engage with information that aligns with their preexisting beliefs (Klapper 1960; Redlawsk 2002; Taber and Lodge 2006). This happens because confronting conflicting ideas creates psychological discomfort, leading individuals to become more selective in the information they consume (Festinger 1957). People accept "confirming" evidence at face value while subjecting "disconfirming" evidence to critical evaluation (Lord et al. 1979). This results in people gravitating toward information that aligns with (confirms) their existing views, a phenomenon widely studied in the context of echo chambers and filter bubbles (Pariser 2011; Cinelli et al. 2021; Nguyen 2020). Echo Chambers are environments in which a person encounters only beliefs or opinions that coincide with their own so that their existing views are reinforced, and alternative ideas are not considered.

1.1.3 Need for Computational Frameworks for Analysis

Understanding online news and user responses is essential for developing solutions that address its adverse effects, such as polarization. Although much of this challenge intersects with behavioral and social sciences, it also demands a technical solution due to the influential role online news and social media play in shaping public discourse. For instance, online content recommendation systems dictate the information bubble each user experiences – a phenomenon that, although concerning, is almost inevitable. This is because, with the exponential growth of information, a system that ranks all content equally would be impractical to use, making filtering mechanisms necessary. However, it is practically impossible to design a system free of biases.

Most recommendation systems are designed to optimize for metrics like user satisfaction and engagement, which, while reasonable, can have unintended adverse effects. For example, Facebook's content recommendation algorithms, optimized for engagement rather than content quality, have been linked to the spread of ethnic conflicts, such as the Rohingya conflict (Ansar and Khaled 2023; Stein Tønnesson and Aung 2022; Tähtinen 2024). Much like newsroom editors, content recommendation algorithms determine which stories gain visibility and reach a wider audience, consequently shaping public discourse. Yet, despite their substantial influence, they operate with minimal accountability, prompting serious concerns about transparency and oversight.

Moreover, designing an effective recommendation system is non trivial due to several inherent challenges. For example, a system that prioritizes user satisfaction by suggesting content similar to users' previously liked content or content favored by others with similar preferences may succeed in keeping users satisfied but risks reinforcing echo chambers by limiting exposure to diverse perspectives. In contrast, broadening a user's information bubble by introducing unfamiliar content may seem beneficial in theory, but in practice, not all users will welcome such a cross-cutting exposure as it could be discomforting.

By exploring how online news spreads and how users engage with it on social media, we can develop computational solutions that enhance mass communication, making news dissemination more effective, transparent, and responsible. Exploring the unintended consequences of content recommendation systems, which often prioritize engagement over long-term user well-being, can help identify mechanisms that reinforce bias and misinformation. Ultimately, these insights can help in creating computational frameworks that promote a more inclusive, balanced, and ethical digital ecosystem.

1.2 Thesis Statement

Computational news analysis reveals slants in political news and a moral divide on social media among audiences of different news sources as well as how news coverage by one publisher influences another.

This research integrates social science theories, such as moral foundation theory (Haidt and Joseph 2004; Graham et al. 2009) and war and peace journalism (Galtung and Ruge 1965; Galtung 1986), to develop computational frameworks for automating news analysis, with a particular focus on political and conflict-related reporting, and examines how news content influences social media users by analyzing their responses and engagement patterns.

It applies social judgment theory (Sherif and Hovland 1961) to model user behavior on social networks, assessing how selective exposure to congenial (like-minded) content shapes information-sharing dynamics. Through a multi-agent social simulation, it explores the effects of selective exposure, providing deeper insights into polarization dynamics and the impact of congenial (like-minded) content consumption in social networks.

1.3 Research Questions

Although the analysis presented in this research addresses many relevant research questions, the primary focus will be on the following key questions.

1.3.1 Political News and Reader Reactions on Social Media

RQ_{1a} *Do news publishers contain political slant in election-related news reporting?*
RQ_{1b} *Do user responses to election-related news on social media differ across news sources?*

We analyze news coverage of the 2020 U.S. presidential election across six US news publishers grouped by political leaning (left, right, and center). Publishers are grouped based on political ratings from AllSides (2023). Allsides have been used by several prior works on analyzing political ideology and moral framing in news articles (Baly et al. 2020; Chen et al. 2020b; Mokhberian et al. 2020).

To assess political slant, we examine the favorability of news coverage toward political figures. Particularly, we analyze the sentiments expressed in news headlines toward the two major party's presidential candidates using entity-centric sentiment analysis (Hamborg and Donnay 2021; Hamborg et al. 2021).

Additionally, we analyze user responses to election-related news on social media (Twitter) through a moral lens. We use moral foundation theory (MFT) (Haidt and Joseph 2004; Graham et al. 2009) to identify the differences in user responses and compare how these variations manifest across audiences of different news sources.

We conduct statistical significance tests to quantify the differences across news publishers and user responses.

1.3.2 Moral Framing in Conflict News Coverage

$\mathbf{RQ_{2a}}$ *Does the moral framing of the same conflict-related news events vary across different news publishers?*
$\mathbf{RQ_{2b}}$ *Does the news from one publisher influence the news from another in conflict-related news coverage?*

We compare moral framing in news coverage of the ongoing conflict in Gaza across three major news publishers: Fox News (US), BBC (UK), and Aljazeera (Qatar). Moral framing is assessed based on the moral foundations in news headlines.

To identify the moral framings, we adopt a vector subspace projection approach (Bolukbasi et al. 2016; Field and Tsvetkov 2019). We identify a vector subspace for each moral foundation and determine the moral framing in news headlines by projecting its contextual embeddings onto the identified vector subspaces.

To uncover differences in news coverage across publishers, we compare the moral framing in which the conflicted entities (Israel, Palestine, and Hamas) appear across news publishers. To quantify the differences, we conduct statistical significance tests. Additionally, we use Granger's causality test (Granger 1969) to examine whether one publisher's moral framing influences another by identifying causal associations in the moral framing across news publishers.

1.3.3 War and Peace Journalism

RQ$_{3a}$ *Does the victim and villain portrayal differ in war coverage across news publishers?*
RQ$_{3b}$ *Does the use of war and peace frame differ in war coverage across news publishers?*

We use war and peace journalism framework (Galtung 1986, 1998) to conduct an entity-centric analysis of the ongoing Gaza conflict and compare the portrayal of the conflicted entities across news publishers.

For this, we crowdsource annotations for war and peace frames in news headlines and identify the victim and villain. We use the annotated dataset to fine-tune language models and analyze war news from three major news publishers covering this conflict. We use timeline analysis and statistical significance tests to quantify differences across news publishers.

1.3.4 Effects of Selective Exposure on Polarization

RQ$_{4a}$ *Does selective exposure to congenial information in a social network contribute to polarization?*
RQ$_{4b}$ *Does higher tolerance among users in a social network help mitigate polarization?*

We design and implement a social simulation testbed to understand how information sharing affects polarization dynamics in a social network. Specifically, we analyze the effects of selective exposure (to congenial content) and user tolerance (toward noncongenial content).

We simulate agents (representing users) embedded in a social network, where they interact by sharing posts and providing sanctions (analogous to likes and comments) to received content. We operationalize user behavior using social judgment theory (Sherif and Hovland 1961) and mimic selective exposure by filtering content based on user preferences.

Agents seek to maximize positive sanctions (i.e., approval for their shared content) and gradually adjust their preferences over time to enhance social acceptance. We use the simulation to study how selective exposure to congenial (like-minded) content and varying levels of user tolerance affect users over time. To evaluate changes in the network, we design metrics to measure polarization, network homophily, and user satisfaction.

A Note on News Curation

We curate two datasets, *NewsSlant* and *GazaNews*, for analysis. NewsSlant contains news related to the 2020 US presidential election from six US news publishers. It includes headlines from traditional online news platforms, their content on Twitter, and user reactions on Twitter. GazaNews, contains news from three news publishers, and includes headlines published on their traditional online news platforms.

We focus our analysis on news headlines and not the entire news article. This is because headlines provide readers with the essential context for interpreting the news (Dor 2003), and are critical in shaping how a news story is understood. Readers often devote more time to scanning headlines than reading full news articles (Dor 2003). News stories often originate from the same sources and contain similar content. Publishers use headlines to frame the narrative in a way that aligns with their desired perspective (Weatherly et al. 2007), leading to different interpretations even when the full article is the same (Tannenbaum 1953; Konnikova 2014). Moreover, headlines are crafted creatively to capture attention (Ifantidou 2009), and are typically written by editors, even for articles authored by other journalists (Bell 1991). Thus reflecting the publisher's position better than the full news articles.

1.4 Contributions

We present computational approaches to analyze news and its influence on social media users and create resources that would facilitate further research in this domain.

1.4.1 Slant in Political News and Moral Divide on Social Media

We present a computational approach to evaluate slants in political news and user engagement with news on social media. Slants are assessed based on the favorability of coverage toward political figures via entity-centric sentiment analysis (Hamborg and Donnay 2021), and user responses are analyzed using a language model fine-tuned on the moral foundation Twitter corpus (MFTC) (Hoover et al. 2020). We apply statistical significance tests to quantify the differences across news sources and user responses, providing a systematic analysis of slants and variations in political reporting and user responses on social media.

We apply the presented approach to analyze the 2020 US election-related news and social media response to the news. For this, we curate a dataset of election-related news and user responses to news on Twitter. The dataset incorporates publishers with varying ideological leanings (left-leaning, right-leaning, and non-partisan), to enable comparisons across the political spectrum.

Our findings indicate that election-related news coverage exhibits political slants in both headlines and social media (Twitter), with effects more pronounced on social media. Moral foundations in user responses to similar news vary for different news sources, suggesting a potential moral divide among audiences of news sources with different political leanings.

The curated dataset enables the study of political slants in online news and can be used to enhance our understanding of user engagement with news on social media.

1.4.2 Identifying Moral Framing via Vector Subspace Projections

We present a computational approach to analyze moral framing in conflict-related news. For this, we adopt a semi-supervised vector subspace projection approach (Bolukbasi et al. 2016; Field and Tsvetkov 2019) and identify the moral vector subspaces corresponding to each moral foundation in MFT. The approach relies on lexicons to identify word pairs with opposing moral semantics (i.e., orthogonal in contextual representation vector space), making it efficient and scalable. Moral framing is extracted by projecting contextual word representations onto the identified moral subspace.

We apply this approach to analyze the news on the ongoing conflict in Gaza. For this, we curate a dataset of news on the ongoing conflict from three major news publishers, Fox News (US), BBC (UK), and Aljazeera (Qatar). We conduct an entity-centric analysis of the conflicted entities and compare the moral framing that they appear in across news publishers. We use statistical significance testing to quantify the differences. Additionally, we apply Grangers' causality test (Granger 1969) to identify causal associations between moral framings across news publishers and establish cross-publisher influence.

Our findings reveal significant variations in moral framings across news publishers. News publishers report on different conflicted entities in significantly different moral framings while covering the same news events. A notable trend in our findings is the causal association between moral framings across news publishers, which suggests that moral framing in one news publisher influences moral framing in other news publishers. Causal associations in news coverage suggest a potential agenda-setting, where one publisher controls the narrative by setting the agenda, and others follow.

1.4.3 Analyzing Victim and Villain via War and Peace Journalism Framework

We present a computational model to analyze conflict-related news using war and peace journalism framework. For this, we create a crowdsourced annotated dataset manually annotated for war and peace frames and major conflicted entities annotated as victims and villains in news headlines of the ongoing conflict in Gaza. We fine-tune pretrained language models on the annotated data to create computational models that can differentiate between war and peace frames and victim and villain portrayals in war news.

Our findings reveal significant variations in war and peace frames across news publishers. However, all publishers use both frames. The portrayal of victims and villains shows significant variations across publishers for different conflicted entities. Moreover, these portrayal aligns with each publisher's presummed (based on prior works) political leaning and historical coverage, suggesting potentially partisan reporting. Additionally, we identify practical challenges and pitfalls in applying war and peace journalism framework to analyze conflict news reporting.

The annotated dataset can be used to further the research on analyzing news coverage of conflicts.

1.4.4 Evaluating Selective Exposure via Multiagent Social Simulation

We design and implement a multiagent social simulation to examine how individual behaviors and interactions contribute to the emergence and escalation of polarization through information sharing in a social network. We draw on social judgment theory to model attitude shifts and user behavior and incorporate empirical insights from previous studies. The simulation allows for variations in user tolerance and selective exposure to congenial (like-minded) content, enabling investigation into their effects over time. We define metrics such as user satisfaction, polarization, and network homophily to measure the changes in the network over time and demonstrate the impact of varying user tolerance and selective exposure on the dynamics of polarization within a social network.

Our findings reveal that algorithms tailored to user preferences enhance user satisfaction but inadvertently deepen ideological divides, driving network polarization. Selective exposure to congenial content amplifies polarization, limits content reach, and reduces engagement. Whereas higher user tolerance slows down polarization. However, it leads to lower user satisfaction than when users are selectively exposed to congenial content.

This research contributes to the broader discourse on the implications of social media in public discourse and political engagement, offering a computational framework that can be utilized for further research in understanding and addressing polarization in social networks. The insights from this research are vital for informing social media platforms and content moderation strategies, emphasizing the need for a balanced approach to content recommendation that fosters diverse perspectives to mitigate polarization in online communities.

1.5 Organization

This dissertation is organized as follows. Chapter 2 introduces an entity-centric approach to identifying slants in political news and applies moral foundation theory to analyze user responses to news on Twitter. It includes our findings from the analysis of election-related news and discusses the moral divide between users on social media. Chapter 3 presents a vector subspace projection-based approach to identify moral framing in war news coverage. It discusses our findings from the analysis of the news on the ongoing war in Gaza, showing significant differences in the moral framing across publishers and revealing cross-publisher influence. Chapter 4 introduces the crowdsourced annotated dataset on war and peace journalism and presents an entity-centric analysis of the conflicted entities. Additionally, it highlights some practical limitations of the war and peace journalism framework. Chapter 5 discusses the design and implementation details of the social simulation to investigate how information sharing in a social network affects polarization dynamics. It evaluates factors such as selective exposure and user tolerance and presents findings for the changes in network homophily and user satisfaction over time. Chapter 6 summarizes the important findings and discusses research significance and social impact along with potential future directions.

CHAPTER

2

NEWSSLANT: ANALYZING SLANTS IN POLITICAL NEWS AND USER ENGAGEMENT ON SOCIAL MEDIA

2.1 Introduction

One of the most common ways people, particularly young adults, get political news is via social media (Infield 2020). While social media helps in the quick, large-scale dissemination of news, it also witnesses trolling and hate speech. The polarizing effects of political news can be observed on social media platforms (Hong and Kim 2016). Further, anger is the most common emotion in response to news on social media about politics in a crisis situation across cultures (Oliveira et al. 2022, 2023), highlighting the need for more effective ways of disseminating political news.

People show partisan preference in online news consumption and more often subscribe to news that confirms their existing beliefs (Flaxman et al. 2016). Previous studies suggest that exposure to belief-conforming political information correlates with polarizing people's opinions to align with the political party's values they support (Stroud 2010; Garrett et al. 2014; Kim 2015; Westerwick et al. 2017). News coverage in recent times has exhibited a noticeable increase in the use of polarizing language, especially when mentioning political figures (Chinn et al. 2020), raising concerns about the potential aggravation of existing divisions on contentious social and political issues.

To understand political slants in news and their influence on readers, we analyze the 2020 US presidential election-related news and readers' reactions to political news on Twitter. Using an entity-centric sentiment approach, we identify political slants based on the favorability of news toward the two major parties' presidential candidates. We further identify the news topic to infer the relevant social and political issues being reported. Combining news topics and sentiment content provides useful insights into how public opinion varies (Albanese et al. 2020). Additionally, we identify moral foundations in readers' reactions to the news on Twitter using Moral Foundation Theory (MFT) (Haidt and Graham 2007).

For this analysis, we pick six US news publishers and group them based on political ratings from AllSides (2023). Allsides has been used by many prior works for analyzing political ideology and unfairness in news articles (Baly et al. 2020; Chen et al. 2020b; Mokhberian et al. 2020). The news publisher groups are LEFT (left-leaning), RIGHT (right-leaning), and BALANCED (nonpartisan). To ensure a fair comparison, we pick two LEFT, two RIGHT, and two BALANCED news publishers.

In particular, we will try to answer the following research questions.

RQ_1 Does election-related news from news sources with different political leanings contain different political slants?

RQ_2 Do user responses to election-related news on social media differ across news sources with different political leanings?

To answer RQ_1, we compare political slants across news publisher groups (LEFT, RIGHT, and CENTER). Additionally, we compare the news published on their traditional online platforms with the news published on Twitter. To answer RQ_2, we identify moral foundations in user responses to news on Twitter, and compare across news publishers with different political leaning. We use statistical significance tests to quantify the differences.

We find that election-related news from news sources with different political leanings shows different political slants. LEFT favors Biden (presidential candidate for the left-leaning party), and RIGHT favors Trump (presidential candidate for the right-leaning party). News from BALANCED is less slanted than LEFT or RIGHT; however, shows some slant in favor of the LEFT. The difference in sentiments (towards Biden and Trump) between LEFT and RIGHT is significant. The effects are higher in the news on social media than in traditional online news, suggesting more variance in political slants in the news on social media. Further, news on social media is more sentimental than on traditional news platforms. The increase in the political slant in the news on social media (versus traditional news platforms) is better aligned with the political leaning for RIGHT than LEFT, suggesting RIGHT uses social media to advance its slant more than LEFT.

Moral foundations in readers' reactions to LEFT and RIGHT differ. The differences are statistically significant; however, the effects are very small. Further, the shift in moral foundations (from the mean) differs between LEFT and RIGHT across social and political issues. User engagement (number of reader reactions per tweet) is highest in reactions to the RIGHT and lowest in reactions to BALANCED.

To the best of our knowledge, this is the first work that analyzes how moral foundations differ between readers' reactions to the news with different political leanings across social and political issues on Twitter. Analyzing political slants in news and readers' reactions to such news can aid us in understanding the influence of news in shaping public opinion and help us discern more proficient strategies for news propagation.

Organization. Section 2.2 describes the related works, Section 2.3 describes the dataset, Section 2.4 explains the methodology, Section 2.5 details the results of our analysis, Section 2.6 includes a discussion and underlines the limitations and threats to validity. The paper ends with a conclusion in Section 2.7.

2.2 Background and Related Work

2.2.1 Slanted News and Influence on Readers

Bias exists in the selection and sharing of information, especially news (Hart et al. 2009; Knobloch-Westerwick 2014). Online news consumption shows a partisan preference, with readers spending substantially longer on news sources that align with their political leaning (Garimella et al. 2021). Online news consumers visit a few favorite mainstream news publishers more often than others (Flaxman et al. 2016).

Exposure to attitude-conforming political information correlates with polarizing people's opinions to align with the values of the political party they support (Stroud 2010; Garrett et al. 2014; Kim 2015; Westerwick et al. 2017). Exposure to like-minded partisan news significantly increases political campaign activity, whereas exposure to conflicting news has the opposite effect (Dilliplane 2011). Effects of counter-attitudinal news do not differ from those of balanced news (Wojcieszak et al. 2016). The longer individuals spend on attitude-consistent news from slanted sources, the more immediate attitude reinforcement occurs (Westerwick et al. 2017).

News publishers often have different ideological preferences (Dunaway 2013). Some news publishers align their content to the preference of their readers to ensure better subscription revenue (Gentzkow and Shapiro 2010; Branton and Dunaway 2009), some align their content to attract the audience that their advertisers want (Hamilton 2003). The newsroom's ideology also influences the news content and the political slant in the news (Hassell et al. 2022; Dunaway 2008). News organizations often express their ideological bias not by directly advocating for a preferred political party but by disproportionately criticizing one side (Budak et al. 2016).

Cicchini et al. (2022) study news sharing behavior of Argentinian news media outlets on Twitter and find that media is biased towards the two major national parties and reader groups can be identified based on their news consumption. In the context of the US, prior studies suggest mixed findings. While some suggest strong liberal bias (Groseclose and Milyo 2005), others indicate a centrist stance (Ho et al. 2008; Budak et al. 2016). Garz et al. (2020) find that headlines reported by LEFT are relatively favorable to Democrats, and headlines reported by RIGHT are relatively favorable to Republicans. Interestingly, news framing is not only consequential (Sheshadri and Singh 2019; Sheshadri et al. 2021a) but also differs based on the publishers' ideology (Liu et al. 2019a).

Many prior works have presented methods to identify political slants in news reporting. Groseclose and Milyo (2005) measure the political slant of news publishers by monitoring the relative citation frequency of various policy groups by news publishers and members of Congress. Ho et al. (2008) use positions taken on Supreme Court cases to identify publishers' ideological positions. Gentzkow and Shapiro (2010) measure news media slant based on the similarity of a news publisher's language to that of a congressional Republican or Democrat. Le et al. (2017) measure the slant of news by observing their sharing patterns on Twitter. Budak et al. (2016) measure news media slant based on how positive, negative, or neutral news reports are toward members of different political parties.

Our definition of political slant in the news is inspired by Kahn and Kenney (2002). Kahn and Kenney (2002) identify news slant based on the tone (i.e., positive, neutral, or negative) of news coverage toward incumbent senators. We identify political slants in the news based on news coverage of presidential candidates. We use target-based sentiments toward the two major parties' presidential candidates to infer the favorability of news for different political ideologies and, consequently, political slants in news reporting. An advantage of this approach is that it doesn't require human annotations for political slants or bias. Getting human annotations for large datasets can be expensive. Further, human annotations for political slants in news reports are sensitive to prior knowledge about the news event (Lim et al. 2020) and the differences in sensitivity to bias among annotators (Lim et al. 2018). Moreover, sentiment toward political figures, particularly presidential candidates, is critical as the perception of candidates' traits among voters impacts voters' choices (Markus 1982; Druckman and Parkin 2005).

2.2.2 News and Social Media

Social media is one of the most common ways to get political news (Infield 2020; Elisa Shearer 2019), and influences the level of participation in traditional politics (Bode et al. 2014). Social media platforms can potentially contribute to partisan polarization (Hong and Kim 2016). Politicians use social media for self-promotion, to disseminate information among their followers, and to set the agenda that favors their political interests (Yang et al. 2016). Manifestations of politics can be identified in social media architecture (network structure) and dynamics (information flow) (Nahon 2015). Mainstream news sources and the readers on social media are identifiably partisan (Karamshuk et al. 2016).

Cross-cutting exposure in social networks fosters political tolerance and makes individuals aware of legitimate rationales for oppositional viewpoints (Mutz 2002b). However, algorithmic content filtering, an approach often employed by social media platforms to personalize content recommendations, is unlikely to expose its users to counter-attitudinal news (Levy 2021).

Marozzo and Bessi (2017) analyze how Twitter readers express their voting intentions about a referendum. They use a set of hashtags to categorize each tweet as supporting, neutral, or opposing the referendum. Hashtags are useful in identifying trends on social media; however, hashtags are prone to manipulation (Zhang et al. 2017). In contrast, we use target-based sentiments to determine favorability toward presidential candidates in news tweets and analyze reader reactions based on moral foundations.

Moral Foundation Theory (MFT) (Haidt and Graham 2007) is a social psychological theory that seeks to explain the origins of and variations in human moral reasoning. According to MFT, there are five dimensions of morality, each with two sides—virtue and vice. These five moral foundations are care/harm, fairness/cheating, loyalty/betrayal, authority/subversion, and purity/degradation. Liberals and conservatives rely on different sets of moral foundations; liberals more strongly endorse care/harm and fairness/cheating (i.e., the "individualizing" foundations), whereas conservatives more strongly endorse loyalty/betrayal, authority/subversion, and sanctity/degradation (i.e., the "binding" foundations) (Graham et al. 2009). Further, the usage of moral foundations differs across politicians from different political parties (Roy and Goldwasser 2021).

Roy and Goldwasser (2021) use MFT to identify stance and partisan sentiments of tweets by US parliamentarians and find a strong correlation between moral foundation usage and a politician's nuanced stances. Mokhberian et al. (2020) use MFT to identify framing and ideological bias in the news and find systematic differences across liberal and conservative media. Roy et al. (2021) use MFT to identify moral framing in political tweets and find that moral foundations toward entities differ highly across political ideologies.

Sentiments and topics on social media can be a good proxy for public opinion. Data from social media, such as Twitter, replicate consumer confidence and presidential job approval polls (O'Connor et al. 2010). Twitter user sentiments are more predictive of the upcoming election than mainstream news media opinion polls (Yaqub et al. 2017). We use Twitter data to understand how politically slanted news coverage influences its readers by identifying differences in moral foundations in readers' reactions to the news from news sources with different presumed political leanings.

2.3 Dataset

We present NEWSSLANT, a dataset to analyze political news and its influence on readers. The dataset includes news headlines, news tweets, and reader reactions to news tweets.

We collected news headlines from six US news publishers, covering news stories relevant to the 2020 US presidential elections. To ensure balance in the dataset, we included two left-leaning (CNN and The Washington Post), two right-leaning (Fox News and Breitbart News), and two nonpartisan (balanced) news publishers (USA Today and Business Insider). We obtained the political leaning of news publishers based on ratings from AllSides (2023). Allsides provides political bias ratings to news publishers based on crowd-sourced annotations and expert reviews. Allsides have been used by many prior works, including analyzing political ideology and unfairness in news articles (Baly et al. 2020; Chen et al. 2020b), and identifying moral framing in the news (Mokhberian et al. 2020).

We used Newsapi (2018) to identify URLs for relevant news articles based on a set of keywords (see Table B.1 in the appendix). To scrape news articles from the retrieved URLs, we used Newspaper3k (2014). We collected news articles published between March 25th 2020 (a month before Joe Biden announced his candidacy) and January 20th 2021 (Inauguration Day). We filtered out all the news headlines that didn't mention one of the two major parties' presidential candidates.

In addition to online news, we collected tweets published by the official Twitter handle of the same news publishers for the same period. We used Twitter's developer API (Twitter 2021) to collect the tweets. Additionally, we retrieved all reader reactions (response tweets) to the collected news tweets.

NewsSlant contains ≈36k news headlines and ≈25k news tweets and ≈4M reader reactions (response tweets) to the news tweets on Twitter. Table 2.1 shows the distribution of news headlines, tweets, and reader reactions for each news publisher. The dataset[1] and codebase[2] are publicly available.

[1]https://ieee-dataport.org/documents/newsslant
[2]https://github.com/ahaque2/NewsSlant

Publisher	Leaning	News	Tweets	Reactions
CNN	LEFT	6485	6108	1 704 194
The Washington Post	LEFT	4678	6999	1 051 062
Business Insider	BALANCED	4803	3872	41 731
USA Today	BALANCED	4216	3490	119 377
Fox News	RIGHT	8327	872	648 719
Breitbart News	RIGHT	7377	3243	474 525

Table 2.1: Distribution of news headlines, tweets, and reader reactions.

2.4 Methodology

2.4.1 Analyzing News Headlines and Tweets

Detecting sentiment in the news is challenging as the sentiments expressed are often nuanced and not as explicit as on social media (Hamborg et al. 2021). Further, popular traditional sentiment analysis approaches disregard the aspect for which the sentiment is expressed. This adds challenges when a sentence has mixed sentiments, i.e., positive toward some aspect and negative toward another. To overcome these challenges, we use NewsSentiment (Hamborg and Donnay 2021), a target-based sentiment analysis approach, to identify the sentiments in the news toward the two major parties' presidential candidates.

NewsSentiment uses a bidirectional GRU on top of a language model (LM) and is trained on political news articles. NewsSentiment can identify sentiments toward a specified target in a sentence. For any given sentence, it produces a positive, a negative, and a neutral sentiment score (toward a specified entity) between [0,1], with 0 indicating the lowest and 1 indicating the highest sentiment intensity.

A sentiment analysis approach that works for news text is usually unsuited for tweets. However, news tweets are similar to news headlines in writing style and are more formal than most tweets (i.e., unlikely to have spelling errors or Twitter-specific jargon). Hence, we use the same sentiment detection approach for news headlines and tweets.

We use bootstrapping to compute the confidence intervals and standard errors of the sentiment distributions using Scipy[3].

For a more fine-grained analysis, we identify the topic of the news. We use BERTopic (Grootendorst 2022) to identify the news topics. BERTopic is a transformer-based topic modeling approach that uses BERT to extract meaningful topics from text data. Unlike traditional topic modeling techniques, which rely on matrix factorization and probabilistic models, BERTopic

[3]https://docs.scipy.org

leverages deep learning to better capture the semantic relationships between words.

We preprocess each tweet using the Tweet-preprocessor (Preprocessor 2016). Further, we remove stopwords and unwanted texts common in tweets, such as mentions, URLs, and hashtags. We also remove common words in our dataset, such as party or politician names, that do not relate to any topic (see Table B.2 in the appendix for more details). We use a list of seed words to guide the topic modeling toward more meaningful clusters. To expand the list of seed words, we use a snowball strategy. We first generate topics with an empty seed word list and add seed words based on the top words in the identified topics. We repeat this process thrice, adding more seed words based on the identified topics and regenerating the topics. Each headline is labeled with at most one topic, and we stop after three iterations. We manually merge similar topics; for example, COVID-19 vaccines and COVID-19 cases/death-related news are combined into one topic called COVID-19. Similarly, news on elections relating to mail-in ballots, voter fraud, and polls are combined into one topic, Elections, and so on. The list of seed words and manually merged topics can be found in the Appendix (Table B.3, Table B.4, and Table B.5).

To answer RQ_1 (Does election-related news from news sources with different political leanings contain different political slants?), we investigate differences in how news publishers (across political leaning) report news mentioning the two major parties presidential candidates. We identify *relative* slants in news coverage by comparing the *favorability* of news coverage of the same news stories within and across news publisher groups. Favorability is computed as the ratio of the mean positive sentiment to the mean negative sentiment toward an entity. We compute favorability scores for both candidates separately for each news publisher group. We further analyze the differences in favorability toward each candidate on different topics.

We conduct statistical tests to confirm whether the differences in sentiments between LEFT and RIGHT are significant. To pick a suitable statistical test to compare the distributions, we first identify if the compared distributions are Gaussian (i.e., normal distribution). To verify the normality of the distributions, we use the Shapiro-Wilks normality test (Shapiro and Wilk 1965). Since none of the distributions are normal, we use the nonparametric Kruskal-Wallis H statistical test, which is suitable for nonnormal distributions. We compute the effect size using epsilon square (ϵ^2) (Kelley 1935), which is well suited for the Kruskal-Wallis H test (Tomczak and Tomczak 2014). We interpret ϵ^2 (Table 2.2 in the Appendix) based on interpretation from Field (2013). For all significance tests, we assume the null hypothesis to indicate a similar distribution of sentiments between LEFT and RIGHT and the alternative hypothesis to indicate they are different. We set the significance level, i.e., alpha, as 0.01 to accept or reject the null hypothesis.

	Effect Size	Interpretation
Epsilon-Square (ϵ^2)	[0.00, 0.01)	Negligible
	[0.01, 0.06)	Small
	[0.06, 0.14)	Medium
	[0.14, 1.00]	Large

Table 2.2: Effect size interpretations based on Field (2013).

2.4.2 Analyzing Reader Reactions

We adopt the RoBERTa model (Robustly Optimized BERT Pretraining Approach) (Liu et al. 2019b) to identify moral foundations in reader reactions. RoBERTa is based on Bidirectional Encoder Representations from Transformers (BERT), a transformer-based deep-learning language representation model. While BERT advanced the state-of-the-art for eleven benchmarks NLP tasks, RoBERTa further improved GLUE (Wang et al. 2018a), and SQuAD benchmarks (Rajpurkar et al. 2016, 2018). The RoBERTa model is retrained on ≈58 million tweets to capture the Twitter language specifics and fine-tuned on the Moral Foundation Twitter Corpus (MFTC) (Hoover et al. 2020) to identify moral foundations in reader reactions. The MFTC contains ≈35k tweets annotated for moral foundations based on MFT. Each tweet is annotated with eleven labels (two for each of the five moral foundations and one for the nonmoral foundation). A tweet in the MFT corpus can have more than one label. However, we restrict to one label per tweet, choosing based on the majority label and randomly in case of a tie.

The RoBERTa model, fine-tuned to detect moral foundations, produces a softmax score for each tweet for the ten moral foundations and a score for the nonmoral foundation. Softmax is an exponential function that normalizes the output of a model to a probability distribution over predicted classes that sum up to one. We use the softmax scores as the moral foundation scores for a given tweet.

To answer RQ_2 (Do user responses to election-related news on social media differ across news sources with different political leanings?), we analyze whether readers' reactions to the news on Twitter differ between LEFT and RIGHT. We use the shift in the moral foundation of readers' reactions as a metric for the comparison. Shift in the moral foundation measures how much the readers' reactions differ from the mean. It is computed as the change (in percent) in the moral foundation score from the mean for each moral foundation and is computed separately for each news publisher group. We identify the news topics and the moral foundations for which the shift is substantial.

We further compute user engagement for each news publisher group to identify differences in how engaging each news publisher is on Twitter. User engagement is the average number of reader reactions to each news tweet. We compute user engagement for each topic separately for LEFT, BALANCED, and RIGHT.

2.5 Results

2.5.1 News Headlines and Tweets

Source	Entity	Sentiment	CI_{Low}	CI_{High}	SE
RIGHT	Biden	Positive	0.250	0.261	0.003
		Negative	0.370	0.376	0.002
		Neutral	0.324	0.330	0.002
	Trump	Positive	0.255	0.264	0.002
		Negative	0.391	0.394	0.001
		Neutral	0.326	0.332	0.002
LEFT	Biden	Positive	0.300	0.313	0.003
		Negative	0.275	0.293	0.005
		Neutral	0.318	0.327	0.002
	Trump	Positive	0.202	0.214	0.003
		Negative	0.380	0.384	0.001
		Neutral	0.323	0.329	0.002

Table 2.3: Bootstrapped mean errors. CI: Confidence-Interval and SE: Standard-Error.

We compute bootstrapped standard error and confidence intervals for sentiment distributions toward the two candidates (Table 2.3). The difference between low and high confidence intervals and the standard error is low, indicating that bootstrapped sample means are closely distributed around the actual distribution means, and the sample represents the actual data well. Figure 2.1 compares the sentiment distributions toward the two candidates between LEFT and RIGHT visually.

Table 2.4 shows the favorability scores toward the two candidates in headlines and tweets. Figure 2.2 and Figure 2.3 compare the favorability scores for the two candidates on different topics across news publisher groups. RIGHT has a higher favorability score for Trump for all news topics, and LEFT has a higher favorability score for Biden for all news topics in both, news headlines and tweets. BALANCED is more favorable to Biden than Trump, and in some cases even more favorable than LEFT. LEFT favors Biden and RIGHT favors Trump across all topics.

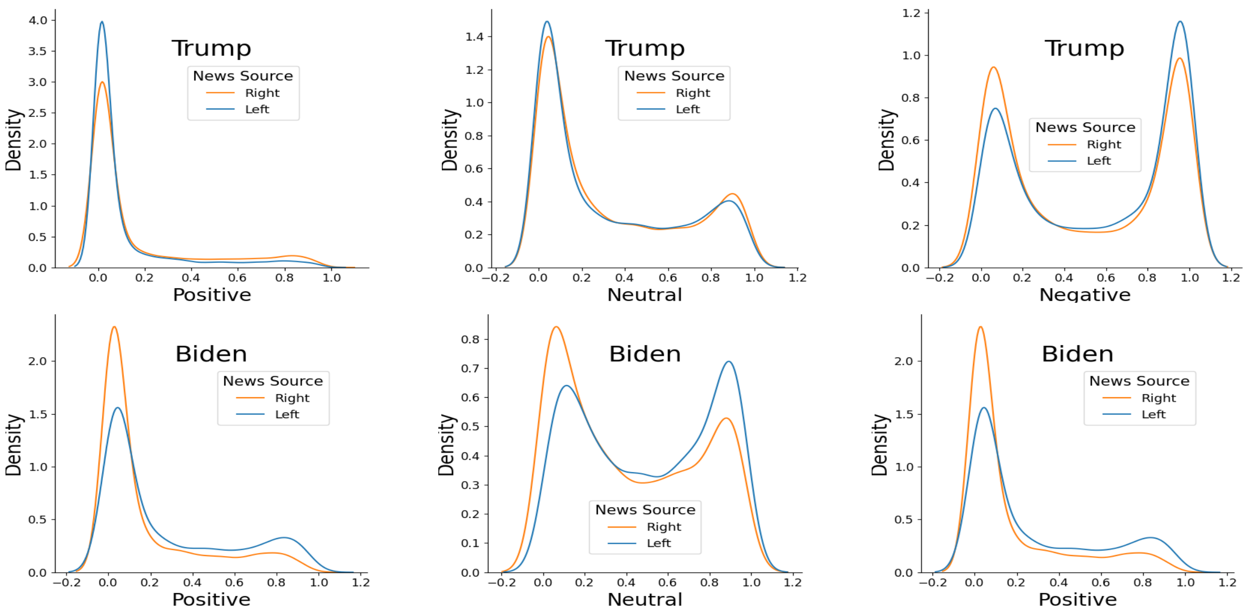


Figure 2.1: Comparing sentiment distributions in headlines between LEFT and RIGHT.

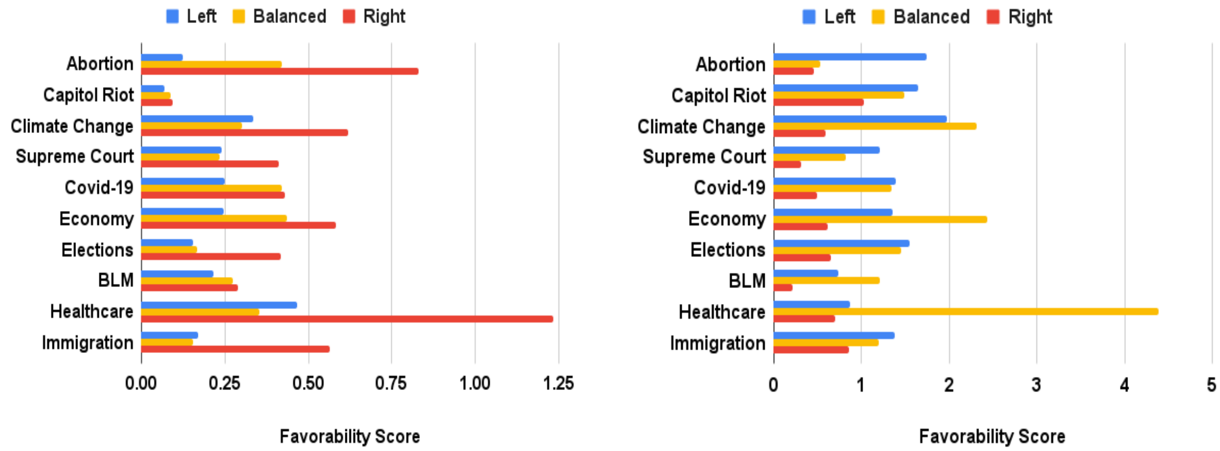
Entity	**Source**	**Source Leaning**		
		LEFT	BALANCED	RIGHT
Biden	Headlines	1.261	1.210	0.435
	Tweets	1.428	1.311	0.393
Trump	Headlines	0.194	0.233	0.322
	Tweets	0.215	0.242	0.494

Table 2.4: Favorability scores across publisher groups.

Favorability is higher for Biden than Trump both in news headlines and tweets.

To isolate the differences between news coverage from LEFT and RIGHT, we conduct statistical significance tests. We compare the sentiment distributions toward the two candidates between LEFT and RIGHT. The differences in sentiment distributions toward the two candidates are statistically significant for both news headlines and tweets (see Table 2.5). However, effects vary, with news tweets showing greater effects than news headlines. For news headlines, the effects are moderate for negative sentiments and small for positive sentiments toward Biden, and the effects are very small for both positive and negative sentiments toward Trump. For news tweets, the effects are moderate toward Biden and small toward Trump for both sentiments.

To conduct a more fine-grained analysis, we identify topics in the news via topic modeling. For news headlines, 79 topics were identified and manually merged into 20 topics. For news tweets, 90 topics were identified and manually merged into 20 topics. Further, we manually



(a) Favorability toward Trump

(b) Favorability toward Biden

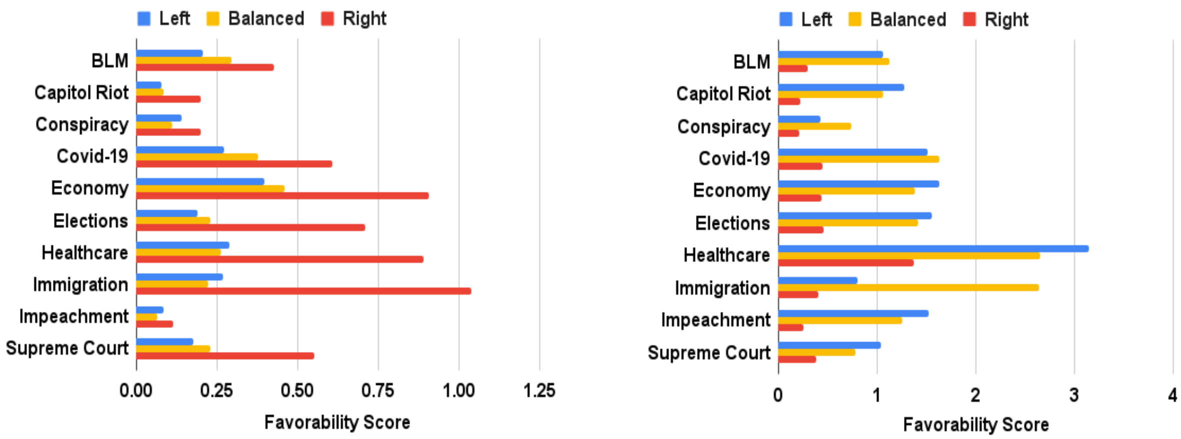
Figure 2.2: Favorability scores of news headlines across news topics and publishers.

Entity	**Sentiment**	**Tweets**		**Headlines**	
		p-value	ϵ^2	p-value	ϵ^2
Biden	Negative	0.00^{*}	0.078	0.00^{*}	0.068
	Positive	0.00^{*}	0.090	0.00^{*}	0.041
Trump	Negative	0.00^{*}	0.036	0.00^{*}	0.007
	Positive	0.00^{*}	0.012	0.00^{*}	0.005

Table 2.5: Statistical test comparing sentiment distributions between LEFT and RIGHT.

identified ten topics (from the identified topics) corresponding to social and political issues in the news. The complete list of topics and subtopics (subtopics combined manually) can be found in the appendix; see Table B.4 and Table B.5.

Most topics discussed are common across news headlines and tweets, with some exceptions. Common topics include Capitol Riots, Climate Change, Supreme Court, Covid-19, Elections, Economy, BLM (Black Lives Matter), Healthcare, and Immigration. Topics exclusive to news headlines are Abortion and Climate Change. Topics exclusive to news tweets are Conspiracy Theory and Impeachment. We further compare sentiment distributions across news topics (Table 2.6 and Table 2.7 show the results). We compare only those news topics for which there are at least ten data points to compare (i.e., a minimum of ten headlines or tweets on the topic for each candidate). For headlines, most topics show a statistically significant difference between LEFT and RIGHT with a few exceptions. News on Healthcare and Capitol Riots don't



(a) Favorability toward Trump

(b) Favorability toward Biden

Figure 2.3: Favorability scores of news on Twitter toward Trump and Biden.

show a statistically significant difference for either candidate for either sentiment. News on BLM and Climate Change show statistically significant differences with moderate effects, but only for Biden. In contrast, news on Immigration shows statistically significant differences with moderate effects but only for Trump. For news on Twitter, topics including Economy and Elections show significant differences with moderate effects for both sentiments and for both candidates. In contrast, news on Immigration and Conspiracy Theory doesn't show a significant difference between LEFT and RIGHT for any sentiment for either candidate. News on Impeachment and BLM shows a significant difference in both sentiments for Biden with large effects, but not for Trump.

Finding 1: News

News sources of the LEFT and RIGHT show signs of political slant in election-related news. The difference in the news coverage of presidential candidates between LEFT and RIGHT is statistically significant, and the effect size varies across social and political issues. The slant is more prominent in the news on Twitter than in traditional online news headlines. The slant on Twitter is more aligned with the political leaning for the news from RIGHT than LEFT.

Topics	Positive Sentiment				Negative Sentiment			
	Biden		Trump		Biden		Trump	
	p-value	ϵ^2	p-value	ϵ^2	p-value	ϵ^2	p-value	ϵ^2
Abortion	0.05	0.04	0.00*	0.12	0.00*	0.09	0.00*	0.10
Healthcare	1.00	0.00	0.03	0.04	0.75	0.00	0.06	0.03
Capitol Riots	0.07	0.01	0.08	0.00	0.03	0.02	0.30	0.00
Supreme court	0.00*	0.06	0.00*	0.02	0.00*	0.07	0.00*	0.02
Economy	0.00*	0.03	0.00*	0.04	0.00*	0.03	0.00*	0.05
BLM	0.00*	0.03	0.09	0.00	0.00*	0.07	0.15	0.00
Climate Change	0.00*	0.07	0.06	0.01	0.00*	0.10	0.26	0.01
Elections	0.00*	0.05	0.00*	0.05	0.00*	0.03	0.00*	0.09
Immigration	0.29	0.01	0.00*	0.06	0.18	0.01	0.00*	0.02
Covid-19	0.00*	0.03	0.00*	0.01	0.00*	0.07	0.00*	0.01

Table 2.6: Statistical test comparing sentiment in news headlines between LEFT and RIGHT.

2.5.2 Reader Reactions

The differences in moral foundations in readers' reactions to LEFT and RIGHT are statistically significant for all moral foundations. However, the effects are very small. 2.8 shows the results.

Table 2.9 shows the shift in moral foundations across news topics (i.e., social and political issues) in readers' reactions to the news from LEFT, and RIGHT. A more detailed result can be found in the appendix (Table B.7). Certain topics fetch more discussion containing moral foundations than the mean for a news publisher group. Topics for which the aggregate moral foundation scores increase across all news publisher groups include Supreme Court, Economy, Capitol Riots, and Impeachment. For discussions related to Elections, Conspiracy Theory, BLM, and Healthcare, the aggregate moral foundation scores decrease in readers' reactions to the RIGHT but increase in readers' reactions to the LEFT. Immigration is the only topic for which the aggregate moral foundation score decreases for the LEFT but increases for the RIGHT and Covid-19 is the only topic for which the aggregate moral foundation score decreases across all news publisher groups.

User engagement differs substantially between LEFT, BALANCED, and RIGHT. BALANCED is the least engaging and the RIGHT is the most. Table 2.10 shows the overall user engagement across different news publishers grouped based on political leaning. User engagement is substantially higher for LEFT and RIGHT than BALANCED. Figure 2.4 compares the user engagement between LEFT and RIGHT across different social and political issues. Few topics have close to equal engagement between LEFT and RIGHT. Topics such as Conspiracy Theory, and Healthcare are more engaging topics for the audience on LEFT (readers responding to LEFT). In contrast,

Topics	Positive Sentiment				Negative Sentiment			
	Biden		Trump		Biden		Trump	
	p-value	ϵ^2	p-value	ϵ^2	p-value	ϵ^2	p-value	ϵ^2
Economy	0.00*	0.12	0.00*	0.02	0.00*	0.06	0.00*	0.03
Conspiracy Theory	0.10	0.02	0.32	0.00	0.01	0.05	0.02	0.01
Elections	0.00*	0.08	0.00*	0.04	0.00*	0.05	0.00*	0.10
Immigration	0.59	0.01	0.06	0.07	0.56	0.01	0.22	0.03
Impeachment	0.00*	0.14	0.86	0.00	0.00*	0.16	0.23	0.01
BLM	0.00*	0.10	0.00*	0.02	0.00*	0.15	0.00*	0.02
Supreme Court	0.03	0.06	0.00*	0.03	0.00*	0.08	0.00*	0.03
Covid-19	0.00*	0.08	0.00*	0.00	0.00*	0.03	0.00*	0.02

Table 2.7: Statistical test comparing sentiment in news tweets between LEFT and RIGHT.

Moral Foundations	p**—value**	effects (ϵ^2)
Care	0.00*	0.0011
Harm	0.00*	0.0004
Fairness	0.00*	0.0004
Cheating	0.00*	0.0002
Authority	0.00*	0.0001
Subversion	0.00*	0.0001
Loyalty	0.00*	0.0000
Betrayal	0.00*	0.0004
Sanctity	0.00*	0.0001
Degradation	0.00*	0.0000

Table 2.8: Statistical test comparing moral foundations between readers' reaction from LEFT and RIGHT.

topics like Impeachment, Supreme Court, Elections, Immigration, Capitol Riots, Covid-19, Economy, and BLM are more engaging for the audience on RIGHT (readers responding to RIGHT). BALANCED has the lowest user engagement for all topics.

2.6 Discussion

News from partisan news publishers shows signs of political slant. This corroborates earlier findings that found systematic differences between liberal and conservative media based on moral framing of the news (Mokhberian et al. 2020), and that political headlines are slanted congenially with respect to the preferences of the news publishers' typical readers (Garz et al.

Topic	Slant	Nonmoral	Care	Harm	Authority	Subversion	Fairness	Cheating	Loyalty	Betrayal	Sanctity	Degradation
BLM	LEFT	**−5**	10	**47**	−2	−3	**20**	−4	2	**22**	**0**	**−3**
	RIGHT	**3**	16	**25**	−7	−12	**−1**	−14	2	**−5**	**85**	**19**
Economy	LEFT	−5	**−5**	−24	1	8	−6	23	1	1	−17	−17
	RIGHT	−13	**22**	−5	14	18	1	24	1	14	−12	−15
Supreme Court	LEFT	−5	−5	−18	18	4	**58**	7	1	−6	14	−3
	RIGHT	−7	−10	−19	23	10	**34**	14	−1	−1	2	−11
Conspiracy Theory	LEFT	**−3**	−19	−12	**−10**	−1	−3	**21**	**−10**	−3	**1**	**1**
	RIGHT	**5**	−5	−9	**20**	0	−14	**−14**	**1**	−9	**−5**	**−4**
Capitol Riots	LEFT	−13	**30**	**40**	9	18	0	−10	6	46	−3	9
	RIGHT	−13	**−3**	**65**	0	16	12	−5	0	55	−10	2
Impeachment	LEFT	−7	−12	**−18**	24	23	4	−7	**0**	**6**	−2	15
	RIGHT	−12	−1	**14**	27	29	0	−9	**33**	**26**	−6	4
Healthcare	LEFT	**−5**	**14**	**4**	8	8	7	**4**	−2	−2	**−6**	**−1**
	RIGHT	**4**	**−1**	**−2**	1	0	12	**−2**	−2	−3	**−26**	**−30**
Immigration	LEFT	**3**	**32**	34	**−9**	**−12**	6	−16	**−1**	**−8**	**10**	**16**
	RIGHT	**−7**	**76**	21	**20**	**16**	10	−13	**25**	**21**	**−17**	**−10**

Table 2.9: Shift in moral foundations. Values in percent (%). ■ Change (> 20%), and ■ Change (>5%) in opposite directions.

	Source Leaning		
	Left	Balanced	Right
User Engagement	210	21	272

Table 2.10: User engagement for news publisher groups.

Finding 2: Reader Reactions

Moral foundations differ significantly between readers' reactions to LEFT and RIGHT. The shift in moral foundations across news topics (i.e., social and political issues) differs between readers' reactions to LEFT and RIGHT. News from the RIGHT is most engaging, followed by the news from LEFT, whereas the news from BALANCED is substantially less engaging than both.

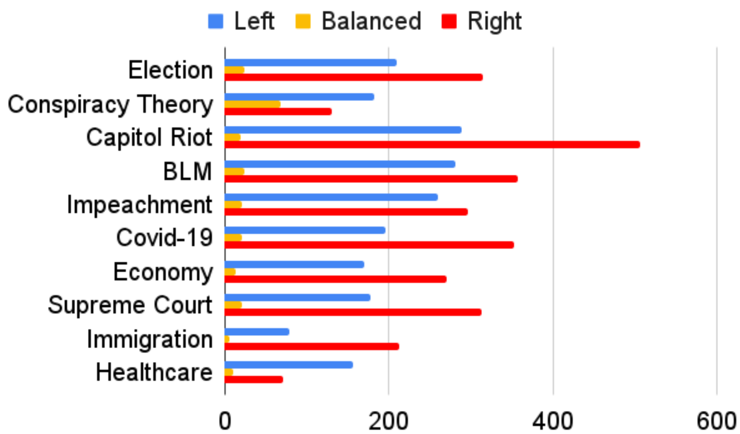


Figure 2.4: User engagement for different news topics.

2020). However, our findings contradict earlier findings that mainstream news outlets in the US present news in a largely nonpartisan manner and do not show favoritism toward either Democrats or Republicans (Budak et al. 2016).

The distribution of sentiments toward the presidential candidates differs significantly between LEFT and RIGHT across traditional online news headlines and news on Twitter. However, the effects are greater for news on Twitter tweets than traditional online news platforms, suggesting more variance in political slant on Twitter. Further, news on Twitter is more sentimental (i.e., higher mean sentiment score) than traditional online news. The increased sentiment in news on Twitter aligns with news publishers' political leaning more for RIGHT than LEFT. News on Twitter from LEFT and BALANCED are more favorable than the headlines for both candidates. However, for news from RIGHT, the favorability in the news on Twitter (compared to news headlines) increases for Trump but decreases for Biden. This suggests that RIGHT is strategically making use of social media platforms to push forward its political ideology.

Biden's favorability is higher than that of Trump across all publishers. This may be because Trump is the incumbent president amidst a global pandemic (Covid-19), and a lot of negative news is mentioned about him. The RIGHT has substantially higher negative sentiments toward Biden in both the news headlines and tweets. In contrast, the negative sentiment toward Trump doesn't vary as much across news publisher groups. Prior research found evidence that news publishers express their ideology not by directly advocating for the preferred political party but by disproportionately criticizing the other party (Budak et al. 2016). The substantial difference in the negative sentiment toward Biden suggests that the RIGHT may be using disproportionate criticism against Biden to advocate relative support for Trump.

Our findings corroborate earlier findings that suggest liberals and conservatives rely on different moral foundations (Graham et al. 2009). The shifts in moral foundations in readers' reactions differ across news topics between LEFT and RIGHT. COVID-19 is the only topic for which the readers' reactions show a consistent shift (i.e., increment or decline) for all moral foundations across all news publisher groups. COVID-19 is also the only topic for which the discussions containing moral foundations decrease across all news publisher groups. Perhaps because many COVID-19 related discussions are about facts and figures, such as symptoms, infection rate, death toll, and so on, and may not contain a moral foundation. Care/Harm and degradation are the only moral foundations that increase in readers' reactions to COVID-19 related news, while all other moral foundations decline. For some topics, the shift is in the opposite direction. This is true for readers' reactions to news on topics such as BLM, Conspiracy Theory, Healthcare, and Immigration.

2.6.1 Threats to Validity

Determining the political slant of a news publisher is a challenging problem. While we take good care of doing a careful analysis to get insights, our methodology has some threats to validity that need to be acknowledged.

First, the presumed political leaning of news publishers is determined based on political bias ratings from AllSides (2023). Though these ratings are generally considered correct and have been used in many prior studies (Baly et al. 2020; Chen et al. 2020b; Mokhberian et al. 2020) to identify political bias in news reporting, these may not be accurate. Further, it is difficult to classify any news publisher as purely left-leaning or right-leaning as they may have mixed stances on different political and social issues. *Second*, we only used two news publishers for each news publisher group. Including more news publishers can potentially change the results. *Third*, we use data from Twitter to understand readers' reactions to the news from LEFT and RIGHT. Though Twitter is a good proxy for public opinion (O'Connor et al. 2010; Yaqub et al. 2017), and has been used by previous studies as a sentinel tool to monitor public opinion (Tavoschi et al. 2020). Data from Twitter can only account for the audience that uses the platform. Further, opinions on Twitter may not necessarily reflect readers' true opinions, and we do not check if the tweets are from real accounts or bots. Though the estimated proportion of bots on Twitter is low, they may play a more vital role in discussions on contentious social and political issues. Thus, any generalization based on the results should be made with caution.

2.6.2 Limitations and Directions

Although we use state-of-the-art models to conduct the analysis, our analysis still has limitations. *First*, we define political slant based on favorable and unfavorable news, which is determined based on the sentiments toward a political entity. While sentiment toward a political entity (over a period of time) could be a good indicator of the political slant of a news publisher, it is far from perfect. *Second*, we look at the two major parties presidential candidates to identify the political slant in the news. However, the news mentions many other political entities that may reveal a different slant. *Third*, topic modeling via BERTopic used to identify news topics (social and political issues) assumes only one topic per news tweet, while a news tweet can potentially discuss more than one topic. All of the above observations suggest important directions for future work. Incorporating changes in framing within topical news (Sheshadri et al. 2021b), and adopting a more nuanced approach to the attribution of blame in political discourse (Xi and Singh 2024a,b) can enhance the efficacy of our methods. We leave this for future work.

2.7 Conclusion

Our results demonstrate that news publishers show signs of political slant in election-related news in traditional online news and on Twitter. News on Twitter is more slanted than traditional online news headlines, and the slant on Twitter is better aligned with the political leaning for the RIGHT than LEFT. Further, moral foundations differ between readers' reactions to the news from LEFT and RIGHT. Consumers of different news publishers often focus on different aspects (moral foundations) of a social or political issue, making it more challenging to reach a consensus or effective conflict resolution.

Algorithmic content filtering, often used on social media platforms to recommend content to readers, could potentially exacerbate political polarization by recommending content that aligns with a user's existing political opinions. The increased use of social media for news consumption and the abundance of choices of news sources make political polarization more likely. Our research highlights the need to identify better ways of disseminating news with reduced polarizing effects.

CHAPTER

3

MORAL FRAMING IN WAR NEWS: A VECTOR SUBSPACE APPROACH

3.1 Introduction

Social media offers immediate access to diverse sources and viewpoints and has transformed how information is consumed and how public opinion is shaped. However, this quick access to real-time news presents challenges in discerning credible, high-quality reporting from misinformation or sensationalized content. This is because news publishers often resort to sensationalizing news (or presentation) to exploit our preference for negative stimuli. Negative emotions are processed more thoroughly than positive emotions (Baumeister et al. 2001) and have a stronger influence (Ito et al. 1998). Consequently, negative information spreads faster than positive information (Rozin and Royzman 2001), and its effects can be seen in online news consumption (Robertson et al. 2023). Emotionally charged negative language in online content correlates with higher user engagement and sharing (Pröllochs et al. 2021b,a; Brady et al. 2017; Berger and Milkman 2012). Easier and faster contagion of online fake news is positively associated with the greater negativity it carries (Chuai and Zhao 2022). Moreover, fake news diffuses significantly faster and farther than real news (Vosoughi et al. 2018).

When sensationalized news laden with emotionally charged language dominates, it distorts public opinion by prioritizing shock value over truth. In particular, framing and presentation of news about conflicts, such as war news, can influence public opinion and, consequently, the outcome of the war. Sensationalizing war news can exacerbate existing conflicts and deepen existing divisions on contentious issues, moving the public discourse away from conflict resolution.

We provide a computational analysis of war news. We evaluate moral framing in war news using moral foundation theory (MFT) (Haidt and Joseph 2004; Haidt and Graham 2007). MFT has been used in earlier works to understand moral framing in political discourse (Graham et al. 2009; Roy and Goldwasser 2021). To evaluate the moral framing of news, we identify vector subspaces corresponding to moral foundations.

We curate a dataset, *GazaNews*, containing news headlines about the ongoing war in Gaza. We choose three popular news publishers (Fox News, BBC, and Aljazeera) based on prior work that has shown differences between these news publishers in reporting news in the larger historical conflict between Palestine and Israel (Amer 2022; Heni and Chandra 2022; Suwarno and Sahayu 2020; Doufesh and Briel 2021). We conduct an entity-centric moral analysis focusing on the conflicted entities, *Israel*, *Palestine*, and *Hamas* (more about the entities in Section 3.3.2). We analyze how the moral framing of the news differ between the three news publishers.

We try to answer the following research questions.

$\mathbf{RQ_1}$. Does the moral framing differ in the news reported by different news publishers for the same entity?

$\mathbf{RQ_2}$. Does the news from one publisher influence the news from another?

To answer RQ_1, we conduct statistical tests to quantify the differences in moral framing between news publishers. We find significant differences between the three new publishers, with effects ranging from small to moderate. News publishers use different moral framing to support their narratives about the same event, suggesting potentially partisan news coverage. We observe the largest effects for Israel, which is presented in a positive moral framing by Fox News but in a negative moral framing by Aljazeera.

To answer RQ_2, we conduct Grangers' causality test (Granger 1969) to identify causal associations (influence) between news from different publishers. In particular, we evaluate if one news publisher's moral framing influences another's. We observe causal associations that suggest agenda-setting. In particular, Fox News coverage (particularly of Israel) exhibits a causal influence on BBC and BBC on Aljazeera, suggesting that Fox News's moral framing of the war shapes the narrative adopted by the other two publishers.

Earlier works on social media discourse on political news reveal a moral divide among audiences of different news sources (Haque and Singh 2024) and a 'basic morality bias,' which suggests that people overestimate blatant immorality approved by political outgroup members

(Puryear et al. 2024). Our analysis extends earlier findings by identifying variations in moral framing across different news publishers. These differences in framing can shape how audiences interpret the same event, potentially reinforcing existing divides on contentious issues. Our findings underscore the need for computational approaches to news analysis that can identify different framing to understand media influence better.

3.2 Background and Related Work

We focus on the news about the ongoing war in Gaza that started in October 2023. Although the Israel-Palestine conflict goes back decades (Gelvin 2021), the ongoing conflict was initiated when Hamas launched a surprise attack on Israel on October 7, 2023, prompting the Israel Defense Forces (IDF) to retaliate by attacking the Gaza Strip. Within weeks, Israeli forces killed disproportionately more Palestinian civilians than the initial Hamas attack that triggered the war. On the one hand, the high number of civilian casualties prompted a lot of criticism and scrutiny of Israel, which is seen as a war criminal by many. On the other hand, Israel is seen as defending itself by many who cite that Israel did not initiate this war. The lack of an obvious aggressor and a victim and additional layers of cultural, religious, and territorial conflicts going back decades between the two involved parties make public consensus on this issue fragile.

Foreign politics has a positive and significant influence on news reports of the conflict in international media (Ukka 2023). Consequently, news publishers adopt different framings when reporting on this conflict, shaping their narratives to align with their respective agendas. News publishers, including The New York Times (USA), Fox News (USA), BBC (UK), and The Times of Israel (Israel), present the conflict from the Israeli perspective, portraying Palestine, especially Hamas, as terrorists and the aggressors in the conflict (Amer 2022; Hajra Chaudhry 2024; Heni and Chandra 2022). Their reporting reinforces Israeli narratives and minimizes Palestinian perspectives. In contrast, news publishers such as Aljazeera (Qatar), Detikcom (Indonesia), and the Russian media portray Israel as the aggressor and Palestine as the victim (Suwarno and Sahayu 2020; Heni and Chandra 2022; Strovsky and Schleifer 2021; Doufesh and Briel 2021).

We evaluate BBC (UK), Fox News (USA), and Aljazeera (Qatar) in this study. In addition to previous research highlighting differences in their coverage of the Palestine-Israel conflict, these three publishers are also classified differently by AllSides (AllSides 2023), a media bias rating organization, based on their political leanings. Fox News is labeled as right-leaning, Al Jazeera as left-leaning, and the BBC as centrist.

3.2.1 Computational Approaches to News Analysis

Several studies have applied machine learning and natural language processing techniques to analyze news. Some popular methods include sentiment analysis (Shapiro et al. 2022; Hamborg 2023; Gu et al. 2024; Hamborg et al. 2021; Hamborg and Donnay 2021; Alonso et al. 2021), emotion analysis (Evans et al. 2023; Oliveira et al. 2023; Aslam et al. 2020), topic modeling (Oliveira et al. 2022; Ghasiya and Okamura 2021; Liu et al. 2020), and event detection (Tong et al. 2020; Nugent et al. 2017; Lin et al. 2024; Qian et al. 2019; Kulkarni et al. 2024; Wang et al. 2018b).

Many approaches use external knowledge and additional resources to analyze news. For instance, Le et al. (2017) measures the slant of news by observing their sharing patterns on Twitter. Budak et al. (2016) measure news media slant based on the portrayal of members of different political parties. Gentzkow and Shapiro (2010) measure news media slant based on the similarity of a news publisher's language to that of a congressional Republicans and Democrats. Ho et al. (2008) use positions taken on Supreme Court cases to identify publishers' ideological positions. Groseclose and Milyo (2005) use relative citation frequency of various policy groups by news publishers and members of Congress to measure the political slant of news publishers. Fan et al. (2019) analyze informational bias by identifying bias-inducing spans in news articles. Lazaridou et al. (2017) analyze political media bias in quotes by analyzing speech verbs used by politicians. Newell et al. (2017) use verifiability of attribution to analyze news.

We extend the computational analysis of news, particularly conflict news, by providing a moral analysis of news using the Moral Foundation Theory (Haidt and Joseph 2004; Haidt and Graham 2007) using a vector subspace projection approach (Bolukbasi et al. 2016; Field and Tsvetkov 2019).

3.2.2 Computational Moral Analysis

Many prior works have used moral foundation theory (MFT) (Haidt and Joseph 2004; Haidt and Graham 2007) to conduct moral analysis of news and social media, particularly to understand moral reasoning in political discourse. Evidence from prior work suggests liberals and conservatives rely on different sets of moral foundations; liberals more strongly endorse the care–harm and fairness–cheating foundations (i.e., the *individualizing* dimensions), whereas conservatives more strongly endorse the loyalty–betrayal, authority–subversion, and sanctity–degradation foundations (i.e., the *binding* dimensions) (Graham et al. 2009). The use of moral foundations differs between politicians from different political parties (Roy and Goldwasser 2021).

Johnson and Goldwasser (2018) created a manually annotated dataset using MFT to annotate Tweets from US parliamentarians and analyze morality in political tweets. Roy and Goldwasser (2021) use this dataset to train a bidirectional LSTM to identify stance and partisan sentiments in tweets by US parliamentarians and find a strong correlation between moral foundation usage and a politician's nuanced stances. Roy et al. (2021) extend this dataset with annotations for moral roles for the entity in each tweet and introduce morality frames. Their findings suggest that moral foundations toward entities differ highly across political ideologies. Hoover et al. (2020) create the Moral Foundation Twitter Corpus (MFTC), which contains tweets annotated for moral foundations. Unlike the dataset proposed by Johnson and Goldwasser (2018), MFTC also contains tweets that are not written by politicians and are not related to politics (e.g., tweets on Hurricane Sandy). Several other works have applied MFT to analyze moral framing in social media content (Haque and Singh 2024; Wang and Liu 2021; Hopp and Weber 2021).

The existing annotated datasets for MFT contain Tweets. Although training a classifier on these annotated MFT datasets effectively identifies moral foundations, its applicability is constrained to the specific domain of the dataset—in this case, Tweets. Due to differences in writing style, the model may not generalize well to news content. To overcome this shortcoming, we employ a semi-supervised approach similar to (Field and Tsvetkov 2019), and use eMFD (enhanced Moral Foundation Dictionary) (Hopp et al. 2021) to identify moral vector subspaces. These subspaces are then used to compute moral foundation scores for news headlines.

The vector subspace approach was proposed by Bolukbasi et al. (2016) to identify and mitigate gender bias in word embeddings. They use gender-specific word pairs, such as *he–she* and *queen–king*, to isolate gender vector subspaces and identify gender bias by comparing the geometric distance of occupation words (such as computer programmer and homemaker) in the identified vector subspace. The approach has been applied to tasks, such as debiasing word embeddings (Dev et al. 2021; Manzini et al. 2019; Dev and Phillips 2019; Shi et al. 2024; Izzidien et al. 2022), semantic similarity (Ishibashi et al. 2024; An et al. 2018), affective analysis (Field and Tsvetkov 2019), word embedding interpretability (Mathew et al. 2020), and moral analysis (Schramowski et al. 2022; Mokhberian et al. 2020).

Our approach differs from previous approaches in important ways. Whereas Mokhberian et al. (2020) use averaging over the difference of word embeddings to compute the subspace, we use principal component analysis (PCA). Moreover, they use static word embeddings like Glove, which work well for word-level semantic similarity but fail to capture contextual information at a sentence level. Instead, we use contextual word embeddings, which are better suited for news headlines. Our work differs from Schramowski et al. (2022) as they identify a moral direction based on pairs of contrastive commonsense morality statements (Hendrycks et al.

2021). Their model identifies a moral subspace with a binary dimension of right and wrong. Whereas we construct a moral vector subspace for each of the five moral foundations in the Moral Foundation Theory (MFT). Our moral vector subspaces are more granular and capture morality as a multidimensional representation rather than a binary (right vs. wrong).

3.3 Dataset

3.3.1 News Curation

We collect news about the ongoing Israel-Hamas conflict using MediaCloud (Roberts et al. 2021). We restrict our analysis to three popular news publishers: *Fox News*, a popular US news publisher, *BBC* (British Broadcasting Corporation) News, and *Aljazeera News*, a Qatar-based news publisher. The selection of these publishers is based on previous studies indicating that they provide varying perspectives on this conflict (Amer 2022; Hajra Chaudhry 2024; Heni and Chandra 2022; Suwarno and Sahayu 2020; Doufesh and Briel 2021). Additionally, media bias ratings from Alsides (AllSides 2023), a media bias rating organization, rate these news publishers as belonging to different ideological leanings. Fox News is rated as right-leaning, Aljazeera as left-leaning, and BBC as centrally aligned.

We gathered 24,815 news headlines published between October 7, 2023 (the day Hamas attacked Israel) and January 19, 2025 (the day a ceasefire was reached). We use manually curated keywords to identify relevant news. Keywords can be found in Table C.1 in the Appendix.

The keyword-based search identifies some irrelevant headlines as well (i.e., headlines unrelated to the Gaza conflict). To refine our dataset, we apply zero-shot learning using Llama-3.2-3B-Instruct (Meta 2024) to filter out irrelevant headlines. Llama-3.2-3B-Instruct is a Large Language Model (LLM) capable of following instructions and producing responses without having previously seen examples of a task, making it suitable for our needs. The prompts we used are in the Appendix.

We end up with 9,656 relevant headlines, 4,321 published by Aljazeera, 3,114 published by Fox, and 2,221 published by BBC.

3.3.2 Conflicted Entity Groups

We identify the entities mentioned in the headlines. For this, we use Spacy[1] to determine the named entities in each headline. We manually grouped the 200 most frequent entities (persons, locations, and organizations) mentioned in the news headlines into four groups:

[1] https://spacy.io/api

ISRAEL, PALESTINE, HAMAS, and OTHERS. Table 3.1 shows the distribution of news published by different news publishers for each entity group.

The entity groups are created in a way that entities in a group either directly or indirectly represent the entity or publicly support it. For instance, the entity group ISRAEL refers to all entities that represent Israel, including references to its people (Israeli), leaders (Netanyahu), and locations (Tel Aviv). This also includes entities that support ISRAEL in this conflict, such as the US (Americans), US politicians (Trump and Biden), and political entities (Pentagon, White House, Republicans).

To ensure accurate grouping, we cross-group entities where appropriate. For instance, some entities may naturally belong to one group but warrant reassignment based on their stance. For example, figures like U.S. Congresswomen Rashida Tlaib and Alexandria Ocasio-Cortez have been openly critical of U.S. support for ISRAEL in this conflict. Hence, instead of grouping them under the entity group ISRAEL, we classify them under the PALESTINE entity group due to their public support for PALESTINE.

Similar to the ISRAEL group, the entity group HAMAS refers to all entities that represent Hamas, which is a Palestinian militant and sociopolitical organization that controls the Gaza Strip (the war zone). The entity group PALESTINE refers to all entities that represent Palestine (but are not directly associated with HAMAS). Entities with no direct connection to any conflicted party and who do not publicly support or oppose either side or whose allegiance is unclear or context-dependent are classified as OTHERS.

Table C.2 in the Appendix shows entity groups.

Source	PALESTINE	HAMAS	ISRAEL	OTHERS
Aljazeera	1372	807	3036	665
BBC	442	558	1310	392
Fox	479	1507	2423	485

Table 3.1: News counts for each entity group.

3.4 Method

We use a vector subspace projection approach (Bolukbasi et al. 2016; Field and Tsvetkov 2019) to identify moral vector subspaces for each of the five moral foundations in the moral foundation

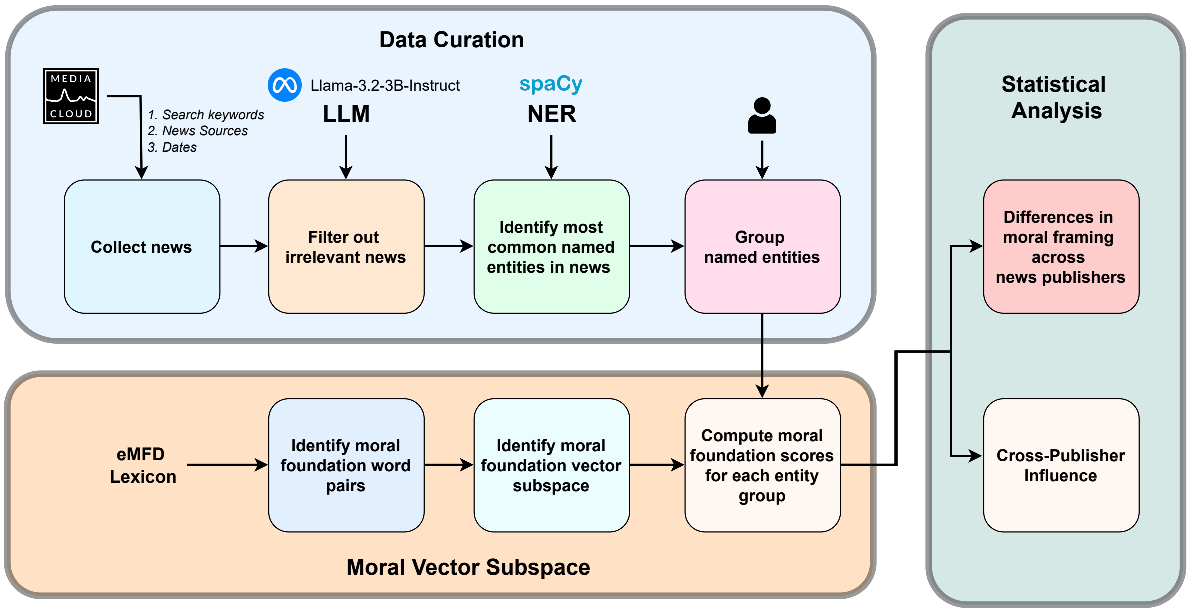


Figure 3.1: Workflow of the methodology.

theory (MFT) (Haidt and Joseph 2004; Haidt and Graham 2007). The approach uses semantically polar opposite word pairs related to a concept to isolate the subspace in vector representations of texts (i.e., word embeddings). For instance, gender-specific word pairs, such as *he–she* and *queen–king*, can be used to isolate gender subspace in the vector space (Bolukbasi et al. 2016). We apply a similar methodology to identify vector subspaces corresponding to each moral foundation. We use the eMFD (extended Moral Foundation Dictionary) (Hopp et al. 2021) to identify relevant word pairs for each moral foundation following an approach similar to Field and Tsvetkov (2019). We compute moral foundation scores for each headline by projecting the contextual embedding of the news headlines to the identified moral vector subspace.

Figure 3.1 shows the workflow of our methodology.

3.4.1 Identifying Word Pairs

We employ a method similar to Field and Tsvetkov (2019) to identify word pairs using the eMFD (extended Moral Foundation Dictionary) (Hopp et al. 2021). The eMFD contains 3270 words annotated for moral foundations via crowd-sourcing. The annotations capture what moral foundation a word is associated with. Additionally, each word has a valence score between −1 and 1 that captures the intensity of the word on the vice–virtue axis, with −1 corresponding to extreme vice and +1 corresponding to extreme virtue.

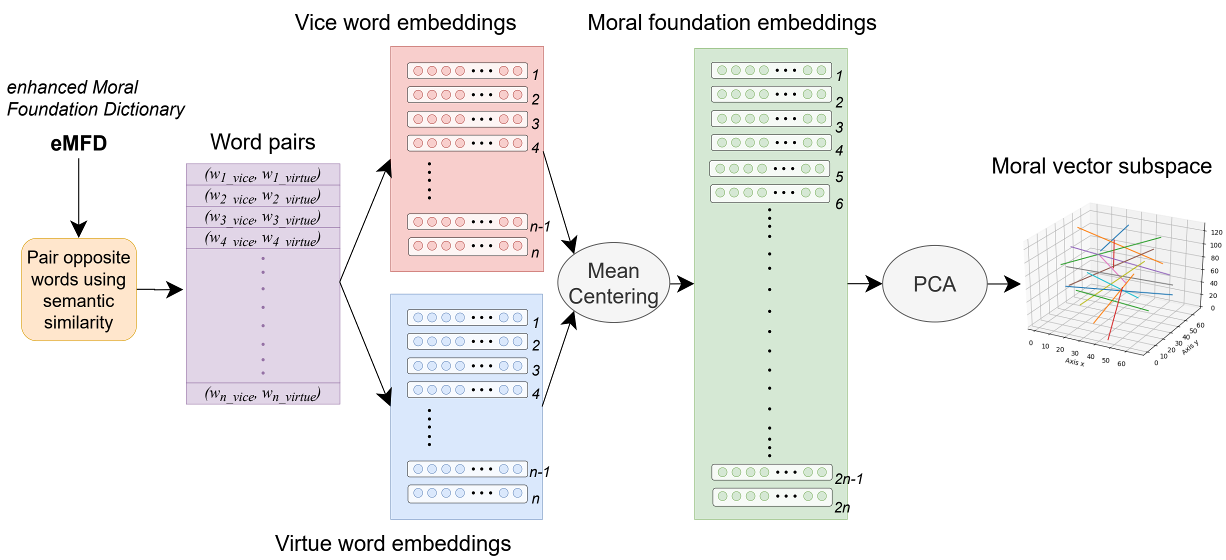


Figure 3.2: Identifying moral vector subspace using word pairs.

We pair words based on their semantic similarity, pairing the most similar words across the highest and lowest-scoring words for each moral foundation. To achieve this, we first identify the top h (high) virtue words and the top l (low) vice words (i.e., words with the most and least scores) for each moral foundation. We then pair the most similar words across the high and low word groups for each moral foundation using cosine similarity, ensuring no repetitions. We then keep the k pairs with the highest semantic similarity. We use grid search to identify the best parameter values for h, l, and k based on the model evaluation (discussed in Section 3.4.3).

3.4.2 Moral Vector Subspace

We use BERT (bert-base-cased) (Devlin et al. 2019) to convert words into their vector representation. BERT is a Large Language Model (LLM) based on Transformers (Vaswani et al. 2017), a deep learning model architecture. BERT produces contextual embeddings for text that can be applied to a variety of downstream language tasks. BERT advanced the state-of-the-art for several benchmark NLP tasks (Wang et al. 2018a; Rajpurkar et al. 2016, 2018; Williams et al. 2018). Moreover, contextual word embeddings generated by BERT encode meaningful affect information (Field and Tsvetkov 2019). We use mean pooling with all BERT layers to get the headline embeddings.

To construct the projection subspace for each moral foundation, we adopt a methodology similar to Field and Tsvetkov (2019). The approach is explained in Figure 3.2. We get the contextual embeddings e_{mh} and e_{ml} for each word in the identified word pairs (w_{mh}, w_{ml}) for a given moral foundation m. We compute their average $\mu = (e_{mh} + e_{ml})/2$. We use the average embeddings to construct a vector subspace (S_m) for the moral foundation m. S_m is constructed by stacking up all $e_{mh} - \mu$ and $e_{ml} - \mu$ embeddings into a $2N \times d$ dimensional vector where N is the number of identified word pairs for each moral foundation.

We apply Principal Component Analysis (PCA) to S_m to identify a projection vector of dimension d that can be used to project contextual embeddings to compute moral foundation scores. PCA is a widely used dimensionality reduction technique in machine learning and statistical analysis. It simplifies complex datasets by applying an orthogonal linear transformation to the data, identifying a set of linearly uncorrelated variables. This process determines the principal components—directions in which the data exhibits the greatest variance. We use the first principle component (i.e., PCA-1) as the moral vector subspace. PCA-1 shows the highest variance and captures most information from the $2N$ word embeddings used to identify the moral vector projection subspace.

	Moral Foundation				
Embedding Model	CARE	FAIRNESS	AUTHORITY	LOYALTY	SANCTITY
BERT	0.533	0.520	0.525	0.540	0.506

Table 3.2: Correlation scores between predicted moral foundations scores and eMFD.

3.4.3 Model Evaluation

To evaluate the performance of our model, we use the scores from the eMFD lexicon (Hopp et al. 2021) as the ground truth moral foundation scores. We report the Pearson correlation coefficient between the moral foundation scores produced by our approach and the ones from the eMFD lexicon. Table 3.2 shows results for each moral foundation.

We use grid search to find the best parameters, i.e., high (h), low (l), and number of word pairs (k) (discussed in Section 3.4.1). Table C.4 in the Appendix lists the best parameters based on the grid search.

3.4.4 Identifying Moral Framing of Headlines

We identify the moral framing of a headline by computing its moral foundation scores. We calculate a score for each moral foundation in every news headline by projecting the contextual

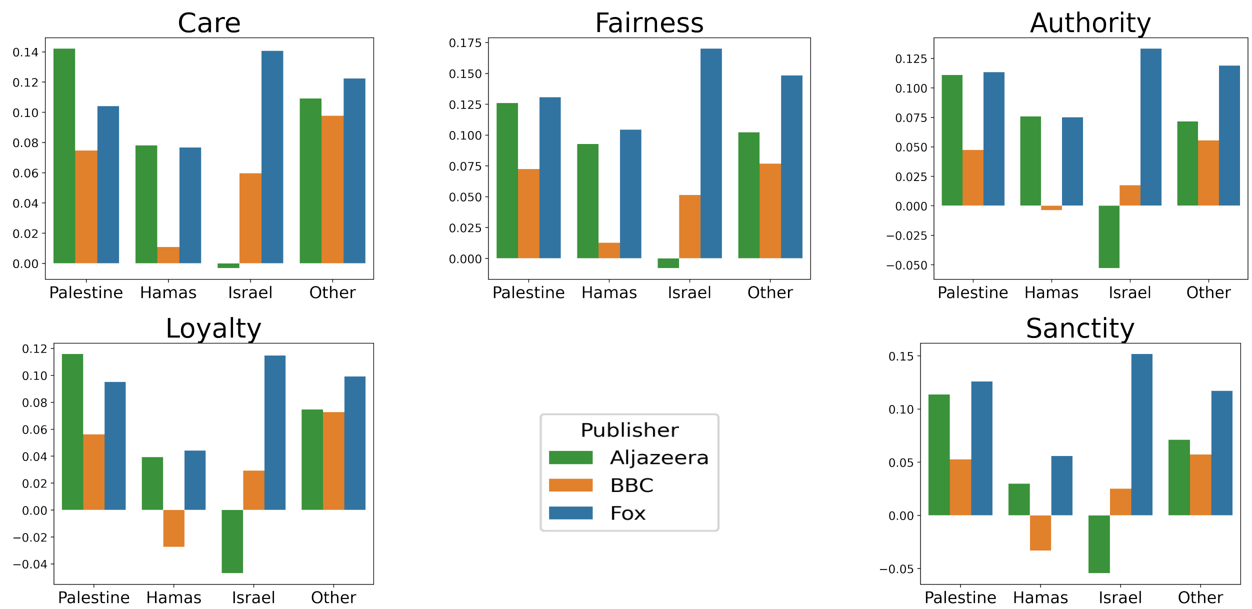


Figure 3.3: Mean moral foundation scores across publishers for different entity groups.

embedding of the headline to the moral foundation projection subspaces. Since we only consider the first principal component, i.e., PCA-1, the projection returns a one-dimension value. This one-dimension value is the moral foundation score for a given news headline. The moral score for each entity group is computed by averaging the moral foundation scores over all entities in a group.

The codebase and dataset for this work is publicly available on Github[2].

3.5 Analysis and Findings

3.5.1 Entity-Centric Moral Framing Analysis

To answer RQ_1, we compare the distributions of moral foundation scores by grouping the headlines based on the entities mentioned in the headline. We group headlines into four entity groups discussed in Section 3.3.2, namely, PALESTINE, HAMAS, ISRAEL, and OTHERS. To ensure robustness in our analysis, we discard headlines with overlapping entity groups.

We compare the distributions for each moral foundation for the same entity group between pairs of news publishers. Additionally, we conduct statistical tests to assess if the differences are statistically significant. To choose the appropriate statistical test, we first conduct a normality test using the Shapiro-Wilks normality test (Shapiro and Wilk 1965). Since the distributions are

[2] https://github.com/ahaque2/Moral_Analysis_of_News

Entity	News Source		Moral Foundation									
	Pub 1	Pub 2	CARE		FAIRNESS		AUTHORITY		LOYALTY		SANCTITY	
			p-val	ES	*p*-val	ES	*p*-val	ES	*p*-val	ES	*p*-val	ES
Palestine	Alj	BBC	1.04e-02	0.242	5.01e-02	0.190	2.02e-02	0.223	2.76e-02	0.211	2.63e-02	0.213
	Fox	BBC	4.62e-01	0.110	1.58e-01	0.208	1.09e-01	0.237	3.34e-01	0.142	7.53e-02	0.263
	Alj	Fox	3.04e-01	0.132	8.98e-01	0.017	9.52e-01	0.008	5.76e-01	0.071	7.43e-01	0.042
Hamas	Alj	BBC	1.94e-02	0.305	**6.16e-03**	0.358	**5.80e-03**	0.361	2.03e-02	0.303	3.46e-02	0.275
	Fox	BBC	1.08e-02	0.269	**5.05e-04**	0.373	**2.24e-03**	0.326	**5.27e-03**	0.298	**8.96e-04**	0.357
	Alj	Fox	9.58e-01	0.005	6.36e-01	0.049	9.75e-01	0.003	8.44e-01	0.020	3.03e-01	0.107
Israel	Alj	BBC	**6.64e-07**	0.252	**8.09e-06**	0.222	**1.42e-07**	0.264	**4.19e-09**	0.298	**5.14e-09**	0.294
	Fox	BBC	**1.37e-08**	0.304	**4.79e-16**	0.442	**4.33e-15**	0.427	**2.90e-09**	0.320	**2.87e-17**	0.461
	Alj	Fox	**1.79e-33**	0.558	**3.44e-47**	0.660	**6.72e-51**	0.690	**4.28e-41**	0.619	**8.56e-60**	0.754
Other	Alj	BBC	4.36e-01	0.044	8.50e-02	0.099	2.76e-01	0.062	8.94e-01	0.008	3.57e-01	0.053
	Fox	BBC	2.39e-01	0.096	**8.14e-04**	0.273	**2.59e-03**	0.244	2.04e-01	0.104	**5.13e-03**	0.228
	Alj	Fox	5.18e-01	0.051	2.61e-02	0.179	2.19e-02	0.181	2.32e-01	0.095	2.82e-02	0.174

Table 3.3: T-test p-values (*p*-val) and effect size (ES) comparing entity-centric moral foundation scores between news publishers. p-values < 0.01 are **bolded**.

approximately normal, we choose the t-test (unpaired) for statistical testing.

To quantify the magnitude of the differences between compared distributions, we compute the effect size for each test. We use Cohen's d (Cohen 1988) to compute the effect size. Cohen's d ranges between 0 and 1. A value of 0.2 represents a small effect, 0.5 represents a medium effect, and 0.8 represents a large effect.

Figure 3.3 compares the mean moral foundation scores for each entity group between news publishers.

Moral foundation scores exhibit the greatest variation in coverage related to ISRAEL. For all moral foundations, Aljazeera demonstrates consistently negative mean values, indicating that ISRAEL is mentioned in headlines associated with vices more than virtues. In contrast, Fox News reports some of the highest mean moral foundation scores for headlines mentioning ISRAEL, suggesting a stronger association with virtues. Similarly, BBC has a positive mean score for ISRAEL for all moral foundations, indicating association with virtues; however, these values are lower than those observed for Fox News.

In coverage related to PALESTINE, Aljazeera exhibits higher mean scores for the moral foundations of care and loyalty than Fox News. However, its mean scores for fairness, authority, and sanctity are slightly lower. BBC, in contrast, reports lower but positive mean scores across all moral foundations compared to both Aljazeera and Fox News.

In coverage related to HAMAS, BBC exhibits negative mean moral foundation scores for loyalty, sanctity, and authority. In contrast, BBC reports positive mean moral foundation scores for care and fairness. When comparing Aljazeera and Fox News, moral foundation scores for care and authority are similar. However, Fox News reports slightly higher scores for fairness, loyalty, and sanctity.

Moral foundation scores for other entities are generally comparable across the three news publishers. However, Aljazeera reports lower mean scores across all moral foundations than BBC and Fox News. Between BBC and Fox News, BBC exhibits higher mean scores for care and loyalty, whereas Fox News reports a higher mean score for authority.

Finding 1: Moral Framing of News

Moral foundations differ across news publishers for the same entity. The differences are particularly pronounced for Israel, with Fox News showing favorable coverage and Aljazeera showing unfavorable coverage. Moral foundations on BBC are neutral, with all mean moral foundation scores for Israel higher than Aljazeera but lower than Fox News.

3.5.2 Statistical Significance Test

News publishers show significant differences in moral framing in war news. Table 3.3 shows the p-values obtained from the t-test comparing the distribution of the moral foundation scores between news publishers for each entity group. Lower p-values indicate a higher likelihood of statistical significance. We use an $\alpha = 0.01$ to determine statistical significance, i.e., p-values lower than 0.01 indicate that the differences between the compared distributions are statistically significant.

Our analysis reveals the highest level of differences in news coverage mentioning ISRAEL across all moral foundations. Moreover, the differences are statistically significant between all pairs of news publishers, indicating that coverage of ISRAEL varies a lot across different news publishers.

For HAMAS, the moral foundations of fairness and authority show statistically significant differences between Aljazeera and BBC. All but the care moral foundation shows significant differences between Fox News and BBC.

Coverage of other entities exhibits statistical significance for fairness, authority, and sanctity between Fox News and BBC. For PALESTINE, no comparison yields a significant difference

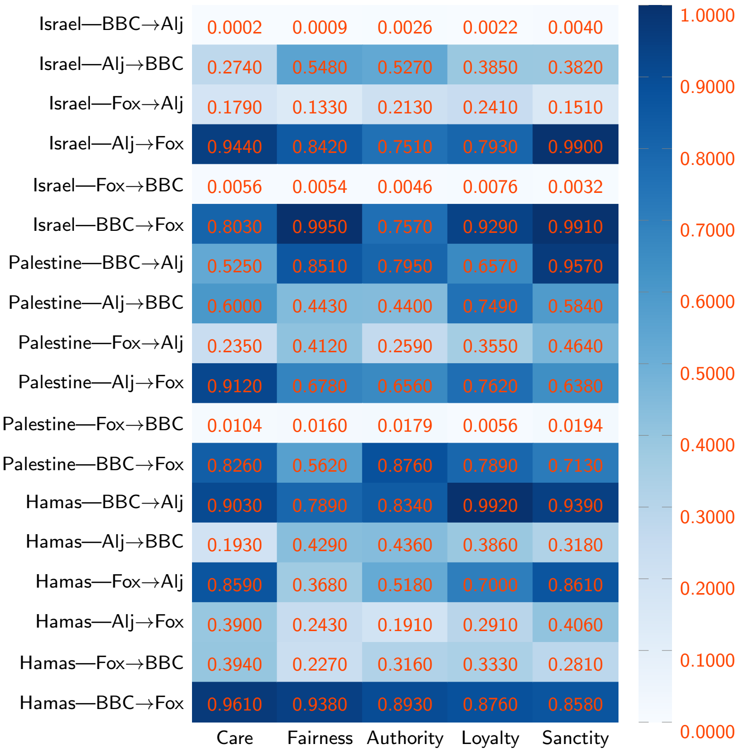


Figure 3.4: Heatmap of cross-publisher influence for the same entity news coverage. Values indicate p-values.

between any publisher for any moral foundation, suggesting a relatively uniform moral framing for PALESTINE across different news publishers.

Table 3.3 includes the effect sizes for each comparison. Consistent with the statistical test results, the largest effects are observed in comparisons involving ISRAEL, ranging from small to moderate, with some values approaching a large effect—such as the sanctity moral foundation between Aljazeera and Fox News. For PALESTINE and OTHERS, effects remain small across all comparisons, with the highest value at 0.273 (for fairness for OTHERS between Fox News and BBC) and the lowest at 0.008 (for loyalty for OTHERS between Aljazeera and BBC). In the coverage of HAMAS, effects are generally larger than those observed for PALESTINE and OTHERS, but lower than those for ISRAEL. Surprisingly, for HAMAS, effects between Aljazeera and Fox News are consistently smaller across all moral foundations than those between Aljazeera and BBC or Fox News and BBC.

3.5.3 Analyzing Cross-Publisher Influence

To answer RQ_2, we use Granger's causality test (Granger 1969), a statistical hypothesis test for determining whether one set of observations is predictive (influences) of another. We employ

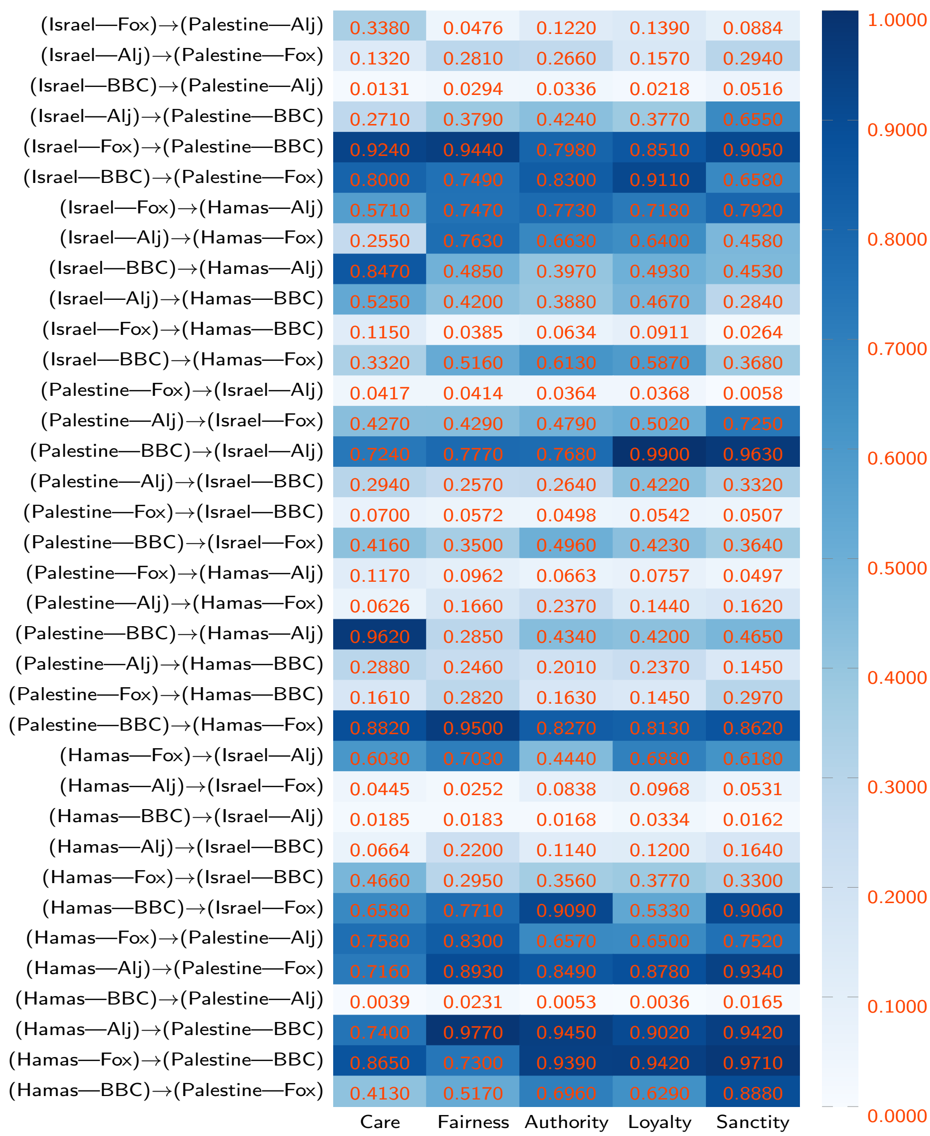


Figure 3.5: Granger Causality p-values for comparisons involving different entities between different publishers. *(Entity1—Pub1)→(Entity2—Pub2)* means influence *of* (Pub1's Entity1's coverage) *on* (Pub2's Entity2's coverage).

Finding 2: Differences Between News Publishers

News publishers show significant differences in moral framing for the same entity with small to moderate effects. The effects are highest for Israel, which shows the most differences, followed by Hamas. The differences in moral foundations for Palestine are relatively low, indicating the coverage differs most for ISRAEL and least for Palestine.

Granger's causality test to evaluate if we can predict the moral foundation scores in headlines from a publisher based on moral foundation scores in the headlines from another publisher.

The Granger causality test is applicable only to stationary time-series data, which means the data should have a constant mean and variance. Nonstationary data may include confounding factors such as seasonal effects and trends. To ensure the data is stationary, we use the Augmented Dickey-Fuller (ADF) Test with a significance level of 0.001 to reject the null hypothesis. All moral foundation score distributions we evaluate are stationary, making Granger's causality test applicable.

We use the F-distribution to conduct Granger's causality test. We compute each entity group's daily mean moral foundation scores. Then, we examine whether one publisher's average daily moral foundation scores can predict (i.e., exhibit a causal association with) another publisher's moral foundations. We apply a lag of one day and conduct the analysis for both the *same entity* and between *different entities*. For instance, we test if the mean moral foundation scores in news mentioning ISRAEL from Aljazeera for a given day are predictive of the mean moral foundation scores for ISRAEL (same entity) and for PALESTINE (a different entity) on Fox News for the following day. We report p-values for all causal association tests. We use statsmodel (2023) to conduct the ADF and Granger's causality test.

Causal associations are directional; a causal relationship from Publisher 1 to Publisher 2 does not imply a reverse association from Publisher 2 to Publisher 1. We compute causal associations for each publisher pair in both directions.

Table 3.4 shows the results of Granger's causality test for the same entity coverage across publishers. Although most comparisons yield a p-value > 0.01, implying no influence, some tests yield a low p-value. The coverage of ISRAEL reveals a causal association between Aljazeera and BBC, where moral framing in the news on BBC for ISRAEL (Granger) causes moral framing in Aljazeera coverage, i.e., Aljazeera's coverage of ISRAEL is influenced by BBC coverage. Similarly, a causal association is observed between BBC and Fox, with Fox (Granger) causing coverage of ISRAEL on BBC. This suggests that it is likely that Fox influences the portrayal of ISRAEL on BBC, and BBC influences its portrayal on Aljazeera. Another entity whose coverage shows a potential cross-publisher influence is PALESTINE. A causal relationship is observed between

BBC and Fox News, with Fox News (Granger) causing BBC's coverage. Other tests yield a high p-value, implying the influence is unlikely.

Results for some of the different entity comparisons yield low p-values implying potential cross-entity influence (Figure 3.5). We observe a significant influence between Aljazeera's coverage of ISRAEL and Fox News's coverage of PALESTINE (with Fox News's coverage of Palestine influencing Aljazeera's coverage of ISRAEL) and between Aljazeera's coverage of PALESTINE and BBC's coverage of HAMAS (with BBC's coverage of HAMAS influencing Aljazeera's coverage of PALESTINE). Additionally, we observe low p-values indicating a potential influence between Aljazeera's coverage of HAMAS and Fox News's coverage of PALESTINE (with Aljazeera influencing Fox News); BBC's coverage of HAMAS and Aljazeera's coverage of ISRAEL (with BBC influencing Aljazeera); and Fox News's coverage of PALESTINE and BBC's coverage of ISRAEL (with Fox News influencing BBC).

Finding 3: Cross-Publisher News Influence

News Publishers show a causal association in moral foundations, which suggests agenda-setting, where one news publisher sets the agenda and others follow. Additionally, we observe cross-entity causal associations between news publishers, which suggests the possibility of reactionary news reporting, where news coverage of some entity by a publisher influences news coverage of another entity by another publisher.

3.6 Discussion

The vector subspace projection approach uses the moral foundation vector subspace and directly extracts moral foundation scores from the contextual text representation. This method requires no training and only needs pairs of opposite words, which represent each moral foundation, to isolate the corresponding vector subspace. We identify these word pairs by pairing semantically opposite words from the eMFD (enhanced Moral Foundation Dictionary).

We chose eMFD as it contains a larger number of annotated words (≈ 3200) than other dictionaries, offering evaluation on a larger set of words. MFD 2.0 extends MFD and contains 210 words per moral foundation but offers no better (or worse) construct validity than the original MFD (Frimer et al. 2019). MoralStrength (Araque et al. 2020) extends MFD and creates a dictionary containing a total of ≈ 1000 lemmas (word roots) along with moral intensities. However, though the moral intensity scores in MoralStrength are crowdsourced, the words themselves are identified from social media posts using semantic similarity with seed words.

Our semi-supervised vector subspace projection approach to compute moral foundation scores for news headlines avoids the challenges of supervised learning (e.g., a regression model) in controlling for confounding factors. For example, many people-oriented words (e.g., hero, dictator) have high authority, while inanimate nouns (e.g., window, bed) have low authority. This could cause a supervised model to focus on the difference between noun classes rather than the actual moral foundations. Additionally, supervised learning could lead to overfitting and poor generalization on unseen data.

Our findings are consistent with earlier works. We observe that Fox News offers favorable coverage of ISRAEL, which aligns with earlier findings (Heni and Chandra 2022). Similarly, we observe BBC provides more positive coverage of ISRAEL than Aljazeera, which is consistent with earlier findings (Amer 2022). In contrast, Aljazeera offers sympathetic coverage of PALESTINE and critical coverage of ISRAEL, which aligns with prior studies (Doufesh and Briel 2021). Although BBC's coverage of ISRAEL is more positive than Aljazeera's, it is still less positive than Fox News, indicating that BBC is more center-aligned compared to the left-leaning Aljazeera and right-leaning Fox News. This is consistent with news bias ratings from AllSides (2023).

Our analysis reveals a concerning trend of cross-publisher news influence. We observe that moral framing in the news from one publisher influences another, suggesting *agenda-setting*, where one news publisher shapes the coverage of others. For example, Fox News influences BBC, and BBC influences Aljazeera (in coverage of ISRAEL). Furthermore, we observe this influence extends across entities. For example, coverage of PALESTINE on Fox News influences Aljazeera's coverage of ISRAEL and HAMAS. Although influence across publishers for the same entity can often be explained by external events or shared reporting sources, cross-entity associations are more complex and less expected. These findings suggest a worrisome pattern of *reactionary reporting*, where a news publisher's coverage of one entity is influenced by how the competing publisher covers another conflicted entity.

3.6.1 Threats to Validity

Although we were cautious in our analysis and employed a rigorous methodology, there are some threats to validity that need to be acknowledged.

First, we use the extended Moral Foundation Dictionary (eMFD) to identify word pairs to isolate moral foundation vector subspaces. While the eMFD lexicon is manually curated through crowdsourcing, moral foundation annotations are inherently subjective, as individual perceptions of morality can vary. Moreover, words can have different meanings and intensities depending on the context in which they are used. As a result, the quality of the lexicon may influence the validity of the identified moral foundation vector subspace and, consequently, the moral framings identified in the news headlines.

Second, we employ semantic similarity in vector space to identify word pairs for each moral foundation using contextual word embeddings. Although this method generally produced meaningful pairs, we observed instances where some pairs, while closely aligned in vector space, were not semantically related. This may have led to the inclusion of some erroneous word pairs. To mitigate this issue, we experimented with incorporating antonym word pairs to ensure that meaningful word pairs were used to identify vector subspace; however, this approach yielded suboptimal results, hence we relied on using semantic similarity using contextual embeddings to identify the word pairs.

Third, although our approach is effective at correctly identifying the moral framing in most cases, it might fail in some cases. This is because the moral foundation scores we compute correspond to the moral context of the headline but do not establish the polarity of the news headlines. A positive moral foundation score isn't necessarily evidence of positive coverage. For instance, a headline with a high score for care may be criticizing the lack of care or appealing for more care. Our approach doesn't effectively differentiate between such cases. We consider the identified moral framing as the moral context in which different entities are reported in the news.

Fourth, we use the Granger causality test to evaluate if news reporting from one publisher influences (causes) another. However, it's important to note that Granger causality doesn't prove direct causation and should be interpreted cautiously. Granger causality means Y helps predict X, but it does not prove that X depends on Y in a causal sense. A key limitation of the Granger causality is that it does not account for potential confounding variables that may influence both the predictor and the outcome. If relevant factors are omitted from the model, the observed relationship may be spurious rather than indicative of true causality.

3.6.2 Limitations

Despite utilizing state-of-the-art models and appropriate resources, our analysis has certain limitations that must be acknowledged.

First, our data collection relies on Media Cloud's keyword-based search, which could omit relevant news or include irrelevant ones. Consequently, this approach introduces a potential selection bias. To mitigate the selection bias, we iteratively improve our keyword list by spot-checking a subset of identified headlines. Additionally, we use a list of exclusion keywords to avoid including headlines related to other ongoing conflicts, such as the Ukraine-Russia conflict, which may otherwise be mistakenly included due to overlapping terminology. We further filter out irrelevant headlines using a large language model (LLM) to enhance the robustness of our dataset.

Second, the quality of our analysis is dependent on the performance of the models used. While we ensure robustness by evaluating the model's performance against a lexicon, we do not conduct a manual evaluation of our model on news headlines. Such analysis could potentially reveal additional limitations. However, manual evaluation is resource-intensive and time-consuming and we leave it to future work.

Third, the scope of our study is restricted to three news publishers, and our analysis is limited to their online news content. As a result, the findings may not be generalizable to a broader range of news sources, and any extrapolations should be made with caution.

3.7 Conclusion

We conduct a moral analysis of news on the ongoing war in Gaza. We apply the Moral Foundation Theory and identify moral vector subspaces using a vector subspace projection approach. The moral foundation subspaces are identified using word pairs from the enhanced Moral Foundation Dictionary (eMFD).

We analyze news from three major news publishers: Fox News (USA), BBC (UK), and Aljazeera (Qatar). Our analysis focuses on the moral framing in the headlines which mention a conflicted entity (i.e., ISRAEL, HAMAS, and PALESTINE).

Our findings reveal significant variations in moral framing between news from different publishers. The effects are largest for ISRAEL, which appears in far more positive moral framing on Fox News than Aljazeera. In contrast, Aljazeera includes ISRAEL in a negative moral framing for all moral foundations. Surprisingly, there isn't much difference in the moral framing for PALESTINE and HAMAS between the two publishers. BBC shows relatively neutral coverage in comparison to the other two publishers, particularly for ISRAEL.

In addition to comparing the moral framing, we conducted a causal analysis using Granger's causality test. We analyze if moral framing from one news can be predictive (Granger-caused) of moral framing in another news publisher using a lag of one day, i.e., how well can we predict the moral framing in the news from a publisher given the moral framing from another publisher for a day before? We evaluate causal association for the same entity, i.e., how does moral framing for an entity influence the moral framing in the news from other publishers, and cross-entity causal association, i.e., does moral framing of one entity influence the framing of another from a different news publisher?

For the same entity analysis, we find that Fox News's coverage exhibits a causal influence on BBC and BBC on Aljazeera, suggesting that Fox News's framing of the war shapes the narratives adopted by the other two publishers. For the cross-entity analysis, we observe some comparisons yield a low p-value, indicating a potential causal association between news publishers across entities. This is suggestive of reactionary news coverage, where news about an entity from a publisher triggers a reactionary news about another (possibly conflicted) entity from some other publisher.

Our analysis reveals concerning trends in news coverage where the same news events are framed differently by different news publishers. This is especially concerning for war news as war news could influence public perception and policies, effectively influencing the outcome of the war. This work is a step in the direction of exploring computational methodologies that can effectively identify such trends. We highlight the need for more work in this domain to ensure fair and impartial news that is focused on serving the public interest rather than pushing an agenda and ensuring high revenues or readers.

CHAPTER

4

WAR OR PEACE JOURNALISM: ANALYZING CONFLICT NEWS

4.1 Introduction

The quality of online news is critical in shaping public discourse. Factual and balanced reporting fosters informed discussions and nuanced perspectives, whereas sensationalized news laden with emotionally charged language distorts public opinion by prioritizing shock value over truth. While objective and impartial news reporting is desirable, it can be challenging to deliver. News reporting can be influenced by financial reasons (Dunaway 2013). Moreover, news publishers often align their content to the preferences of their readers (Gentzkow and Shapiro 2010; Branton and Dunaway 2009), advertisers (Hamilton 2003), and their own ideology (Hassell et al. 2022; Dunaway 2008).

Although deviation from journalistic objectivity results in media bias and increases political polarization (Gerber et al. 2009), news publishers frequently sensationalize news to capitalize on our tendency to be drawn to negativity (Baumeister et al. 2001; Ito et al. 1998). Negativity drives online news consumption (Robertson et al. 2023), and garners more views (Chuai and Zhao 2022). However, journalism driven by sensationalizing news for virality can reinforce stereotypes and fuel public anger, exacerbating conflicts and deepening existing divisions on contentious issues.

The influence of the news is amplified during war because a war (1) is an inherently charged situation rife with events of high emotional power and (2) yields a large quantity of news over a short time. As a result of (2), war news may not be vetted to the standards for other news, raising concerns about its credibility and influence. Moreover, war news makes it challenging to maintain journalistic objectivity, particularly if the journalist's own country is involved (Bläsi 2009).

We use war and peace journalism framework (Galtung and Ruge 1965; Galtung 1986) to evaluate conflict-related news coverage. Galtung (1986) identified two distinctive varieties of journalism for reporting on news related to conflicts: *war* and *peace*. War journalism presents sensationalized news, often from a parochial perspective and without a broader context. It focuses on reporting violence and frames a conflict as between "us" and "them". In contrast, peace journalism focuses on conflict resolution and reports on the adverse effects of war and the suffering of people affected by war. Note that war journalism is different from war news. War news is any news about war, whereas war journalism and peace journalism refer to two styles of news and news presentations about war.

To advance the research on war and peace journalism, we created a crowdsourced annotated dataset of war and peace journalism. We annotated conflict-related news headlines with war and peace journalism frames and identified the villain and victim in news headlines.

Table 4.1 provides examples of war and peace journalism. Headline H1a presents a headline that is sensationalized and aligns with war journalism, whereas H1b presents an objective headline that aligns with peace journalism. Headline H2a focuses on the visible effects of war, such as assault and torture, and aligns with war journalistic style, whereas H2b focuses on the invisible effects of war, such as fear about the future and mental trauma, which aligns with peace journalistic style. Similarly, headline H3a provides a platform for elite voices (aligns with war journalism), and H3b shares common people's stories and provides a voice to common people (aligns with peace journalism). Headline H4a provides an update on the war with a focus on immediate effects, such as death and causality, which fuel the conflict. In contrast, headline H4b provides a broader context of war and actions of non-conflicted entities related to war.

We use this annotated dataset to fine-tune a language model and analyze the news coverage of the ongoing war in Gaza. In particular, we will try to answer the following research questions.

RQ_1 Does the victim and villain portrayal differ in war coverage across news publishers?

RQ_2 Does the use of war and peace frame differ in war coverage across news publishers?

For this analysis, we pick three major news publishers reporting on the Gaza conflict: Fox News (US), BBC (UK), and Aljazeera (Qatar). Our choice of publishers is based on prior work

	News Headline	Style	Justification
H1a	*We'll fight with fingernails' says Israeli PM after US warning*	War	Sensationalized
H1b	*Hamas terrorists open fire at bus stop near Jerusalem, leaving 3 dead, 11 injured*	Peace	Objective
H2a	*Israeli hostage describes graphic sexual assault, beatings, torture in Gaza*	War	Visible effects
H2b	*Palestinian mother fears for her children as she wonders about the future after evacuating Gaza City*	Peace	Invisible effects
H3a	*There is 'absolutely no' cease-fire happening in the Israel-Hamas war: Ophir Falk*	War	Elite source
H3b	*Jewish survivor of Hamas Oct. 7 terror attack is 'still reeling,' pleads for peace in Israel*	Peace	Peoples' story
H4a	*At least 40 killed in Israeli attack on al-Mawasi tent camp, officials say*	War	War context
H4b	*South Africa urges ICJ to order Israel to halt Rafah assault*	Peace	Broader context

Table 4.1: Example news headlines illustrating war and peace journalism.

that has shown differences between these news publishers in reporting news in the larger historical conflict between Palestine and Israel (Amer 2022; Heni and Chandra 2022; Suwarno and Sahayu 2020; Doufesh and Briel 2021).

We find significant differences in the portrayal of victims and villains across news publishers. Fox News portrays Hamas as a villain and Israel as a victim, whereas Aljazeera portrays Israel as a villain and Palestine as a victim. In contrast, BBC provides a more balanced coverage. The use of war and peace frames presents a less conclusive picture than the victim and villain portrayal. All publishers incorporate both frames but in different contexts.

Additionally, our analysis highlights certain limitations of the war and peace journalism framework. We find that the use of a peace frame does not necessarily indicate peace-supportive coverage and additional context is needed to determine that. The use of war or peace frames is often driven by the immediate newsworthiness of news for a publisher's audience rather than a deliberate effort to promote either peace-supportive or war-supportive coverage. However, beyond the selection of war and peace journalism frames, how news is presented is critical in its interpretation.

We contribute a crowdsourced annotated dataset of news headlines to train computational models to distinguish war and peace journalism. Our analysis of news coverage on the ongoing Gaza conflict reveals key differences across publishers and provides valuable insights into conflict reporting. Additionally, we highlight important challenges and limitations of the war and peace journalism framework for future research to address.

4.2 Background and Related Work

Galtung and Ruge (1965) delineate the foundational elements of peace journalism by exploring the factors influencing the flow of international news. Galtung (1986) formalized the concept of peace journalism by laying down expectations from peace-oriented news, such as a fair representation of all voices, providing historical context for events, freedom of the press, and focus on people's problems rather than just elites. Later, Galtung (1998) introduced the term *peace journalism* as an alternative to *war and violence journalism*. He argued that there are two ways of reporting on a conflict, the *low road* (war journalism) and the *high road* (peace journalism). The high road focuses on the conflict and its peaceful transformation by discussing relevant background details essential for understanding the conflict. Whereas, the low road focuses on the meta-conflict that comes after the root conflict. Meta-conflict is created by violence and war, and focuses on the question of who wins. It overlooks the historical context and the critical background information essential to explain the conflict.

Dominant in the media, war journalism is violence-oriented and focuses on the meta-conflict, i.e., the visible effects of violence, without providing adequate context. War journalism presents a conflict as a battle, as a sports arena, or as a gladiator circus. It presents a conflict as between 'us' vs. 'them,' demonizing the adversaries (them) while pretending to speak on 'our' behalf. War journalism is *propaganda-oriented*, for example, relying on official sources while revealing their crimes, but not our crimes; *elite-oriented*, for example, focusing only on our suffering while overlooking their suffering; and *victory-oriented*, for example, seeing peace as a result of victory and reporting only after the conflict starts. In this two-sided form of journalism, one side inevitably wins, and the other must lose. The zero-sum perspective draws upon sports reporting where "winning is not everything, it is the only thing."

In contrast, peace journalism is peace-oriented, focuses on conflict resolution, and sees conflict as a challenge to the world without blaming or demonizing 'others.' It explains conflicts in a broader context and focuses on the less visible impacts of conflict, such as trauma or cultural damage. It is *truth-oriented*, for example, it tries to expose propaganda messages disguised as information; *people-oriented*, for example, represents the suffering of ordinary people by giving voice to vulnerable groups instead of heavily relying on official sources or elites; *solution-oriented*, for example, it sees peace as a result of non-violent initiatives and focuses on conflict resolution. Peace journalism shows similarities with the standards of good journalistic ideals.

Although the foundations for war and peace journalism were laid down by Galtung and Ruge (1965), it was popularized among researchers by the works in the 2000s (McGoldrick and Lynch 2000; Lynch and McGoldrick 2007). Many communications researchers have since used the war and peace journalism framework to analyze news related to national and international conflicts (Gouse et al. 2019). Workneh (2011) apply war and peace journalism framework to analyze US and British newspaper coverage of the Somali conflict. Shinar (2009) analyze the coverage of the 2006 Lebanon war in Canadian and Israeli media using war and peace journalism framework. Several prior works have used war and peace journalism frameworks to analyze Asian conflicts (India, Pakistan, Sri Lanka, Indonesia, and the Philippines) (Lee and Maslog 2006; Lee et al. 2006; Lee 2010). Other applications of war and peace journalism frameworks include exploring news framing in the Pakistan-India conflict in elite US newspapers (Siraj 2008), perspective of Talibanisation in Pakistan (Siraj 2010), evaluating war and peace potential of Pakistani news media (Hussain et al. 2021), and analyzing war in Afghanistan (Ottosen 2010).

War and peace journalism framework has been applied to the conflict between Palestine and Israel. Fahmy and Eakin (2014) analyze the coverage of the Mavi Marmara controversy in Haaretz, NYT, and the Guardian, and find significant differences across news publishers in the use of war and peace framing. Tenenboim-Weinblatt et al. (2016) analyze the conflict narratives in the Israeli news media and find that the establishment actors are found to be the most dominant subgroup in news stories, and violence and military action form the most dominant realm of events.

War and peace journalism has been applied to news coverage beyond the context of war, such as analyzing the news on the trade war (Ha et al. 2020), evaluating conflicting framing in news related to immigration (Kalfeli et al. 2022), understanding racism in the news (McMahon and Chow-White 2011), and differentiating between local news reports and national news on drug war (Lacasse and Forster 2012).

To better understand war and peace journalism, some studies have used surveys to qualitatively assess journalists and their attitudes toward their reporting approaches (Brounéus 2011) and explored audiences' responses to war and peace journalism and their perceptions via surveys (Ibrahim et al. 2013). Bläsi (2009) finds that peace journalism is harder to realize in wartime, especially if the journalist's own country is involved. Particularly, journalists with firsthand experience in the Middle East use fewer peace frames than journalists who have primarily reported from other regions in the world (Neumann and Fahmy 2016).

4.3 Dataset

4.3.1 News Curation

We use the same data that we put together for the moral framing analysis of conflict news described in Chapter 3 (Section 3.3).

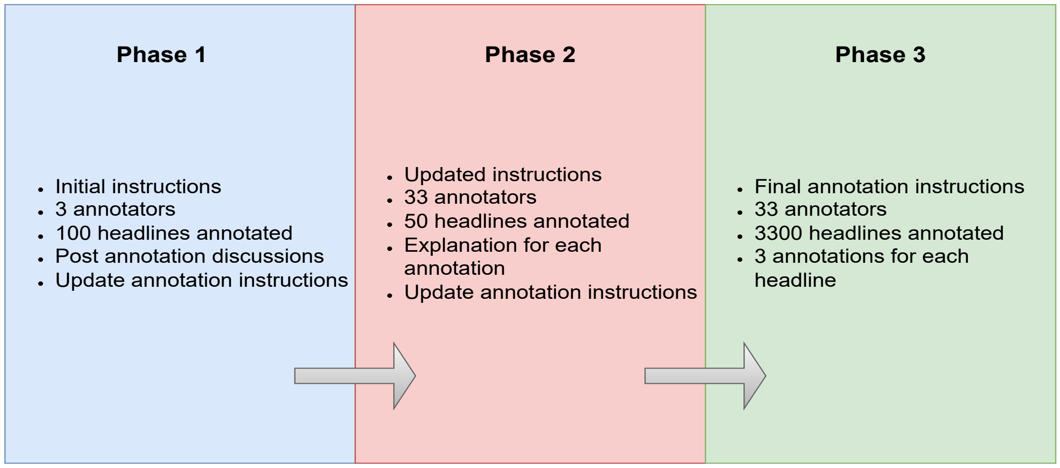


Figure 4.1: Data annotation phases.

We sampled 3300 news headlines to annotate manually. We use a stratified sampling based on month and publisher to ensure a fair representation of headlines from each month and all publishers. The headlines sampled for annotation contain headlines until October 2024, when the annotations were conducted. Headlines later than that are not included in the annotations.

4.3.2 Data Annotation

We define six classes to annotate each headline. Four of these classes correspond to war and peace journalism frames. These frames include *emotiveness frame* (sensationalized vs objective), *effect frame* (visible vs invisible effects of war), *source frame* (elite vs people source), and *context frame* (war vs broader context). Additionally, we annotate the victim and villain in each headline, with labels being Palestine, Hamas, Israel, and None. These classes are described in detail in Table 4.2.

We use crowdsourcing to annotate the data. The annotations were conducted in three phases described in Figure 4.1 In phase one, three annotators (PhD students in Computer Science, including one author) annotated 100 headlines based on initial annotation instructions. Disagreements between annotators were resolved based on a group discussion, and the annotation instructions were modified accordingly. For clarity, we included example headlines in the instructions for each label.

In the second phase, 50 headlines were annotated by 33 annotators (crowdworkers), with each annotator annotating the same 50 headlines. They also provided a justification for their annotations. These annotations and justifications were used to enhance the annotation instructions further.

Class (Labels)	**Description**
Emotiveness	Headline uses emotive/sensationalized or objective language.
Sensational	Headlines that present information in an exaggerated, dramatic, or provocative manner. These can take various forms, such as using emotionally charged language, focusing disproportionately on negative or shocking aspects, or labeling entities involved in the conflict as good or bad (i.e., using demonizing or victimizing language). Headlines featuring dramatic or exaggerated quotes are also considered sensational headlines.
Objective	Headlines that avoid using emotive language and present information neutrally present factual information without exaggeration and do not contain demonizing or victimizing language.
Effect	Headline describes an effect (actions and consequence) of war.
Visible	Headlines that include casualties, destruction, and acts of aggression.
Invisible	Headlines that include psychological trauma, social fragmentation, population displacement, cultural erosion, food and health crises, or
None	Headlines that do not describe an effect.
Source	The source of the headline and who the headline is about (i.e., the subject and object). A headline could provide a voice to, the *elite* or the *common* people.
Elite	Headlines that provide a platform for the authorities and elite voices (and stories), including comments made by elite sources and information that is credited to some elite source. We consider all entities referenced with a name as elite.
People	Headlines that provide a voice to common people and describe peoples' stories and opinions. This includes headlines describing personal stories, common people's acts, and comments.
None	Headlines that do not have a source attribution.
Context	The context of the headline.
War	Headlines that focus on the latest developments in the ongoing conflict. These headlines describe the actions or comments of the conflicted entities or the effects of some event on them.
Broader	Headlines that provide a broader context, including historical contexts, highlight the events that triggered the war, and broader (indirect) consequences. This includes headlines that describe the actions (or comments) of nonconflicted entities on the conflict.
None	if the context is unclear.
Role	The conflicted entities are portrayed as villains or victims.
Villain	An entity that is shown in a bad light (i.e., negatively). In the context of war news, consider the aggressor (the one attacking) as a villain.
Victim	An entity that suffers harm, injustice, and adversity. In the context of war news, a victim refers to the conflicted entity that suffers physical, psychological, social, or economic harm as a direct or indirect consequence of conflict.
None	If a victim or a villain is unclear.

Table 4.2: Annotation Classes and Labels.

In phase three, each of the 33 annotators annotated 300 news headlines. Each news headline was independently labeled by three annotators. The final label for each headline was determined through a majority vote. Headlines, where no majority was reached (i.e., when all three annotators assign different labels), were discarded. Additionally, we discarded all annotations from phases one and two to ensure quality.

We use Fleiss Kappa Fleiss (1971) to compute the interrater agreement between the three annotators for each annotated headline. To ensure quality, we discard annotations from annotators with an interrater agreement below 0.25. If an annotator's agreement for a specific class falls below this threshold, all their annotations for that class are removed. Table 4.3 shows the interrater agreement for each class.

Class	**Count**	**Agreement**
Emotiveness Frame	2601	0.42
Effect Frame	2717	0.48
Source Frame	2834	0.51
Context Frame	2535	0.60
Role-Victim	2512	0.57
Role-Villain	2819	0.66

Table 4.3: News Count and Interrater Agreement.

	War and Peace Frames				**Role**	
Model	Emotive	Effect	Source	Context	Victim	Villain
BERT	80.52	69.1	79.69	77.8	72.39	79.27
RoBERTa	83.83	71.55	82.01	77.93	74.61	81.16
ConfliBERT	84.88	61.46	81.66	75.81	72.01	81.74
ModernBERT	79.3	66.84	79.87	75.6	70.16	74.88
GPT-2	79.82	69.93	78.79	67.68	74.02	75.72
BART	84.18	69.79	80.94	77.36	71.61	80.49
T5	77.79	69.52	65.37	56.18	69.52	69.55

Table 4.4: Model Performance (F1-scores).

4.4 Method

We fine-tune pretrained language models to identify war and peace journalism in conflict news. To choose the suitable model, we evaluate the different architectures and their suitability for the task.

Encoder-only models, such as BERT, excel at tasks requiring deep contextual understanding, such as text classification and entity recognition, due to their bidirectional nature. However, they struggle with generative tasks. Decoder-only models, like GPT, are powerful for text generation and question-answering tasks but suffer from poor ability to encode input sequences with deep contextual representation. Moreover, decoder-only models suffer from attention degeneration over long sequences (Fu et al. 2023). The decoder-only LLMs perform worse on semantic understanding tasks than an encoder-only language model that has vastly fewer parameters (Qorib et al. 2024). Encoder-decoder architectures, such as BART, provide strong performance in sequence-to-sequence tasks such as machine translation and summarization, but they are computationally expensive.

We experiment with encoder-only, decoder-only, and encoder-decoder pretrained language models. However, we include more encoder-only models given their (expected) suitability for the task. For encoder-only models, we experiment with variants of BERT such as BERT (bert-base-cased), Roberta (Roberta-base), and ModernBERT (modernbert-base). For decoder-only models, we experiment with GPT-2. For encoder-decoder models, we experiment with BART (Bart-base) and T5 (t5-base). We also experiment with ConfliBERT (conflibert-cased) (Hu et al. 2022), a domain-specific pretrained language model for conflict and political violence.

4.4.1 Model Fine-Tuning

We fine-tune each model separately for each class. We use mini-batch stochastic gradient descent with momentum and weighted decay using AdamW as an optimizer. To address model overfitting, we add a dropout layer with a dropout of 0.15 (i.e., 15%) before the final output layer. Additionally, we use L2-regularization (λ= 1e-5).

To minimize the impact of confounding factors during model training, we mask all entity mentions. This prevents the model from relying on entity mentions to make predictions, which is crucial because our dataset contains repetitions of specific entities. If entity mentions were included, they could introduce bias. For example, if most headlines belonging to a class contain a particular entity, the model might learn to associate that entity with the class. As a result, it could prioritize that class whenever it encounters the entity rather than relying on linguistic patterns. This would reduce the model's ability to perform well on unseen data. Masking entity mentions improves generalization, because our goal is to train a model that differentiates between war and peace journalism headlines based on language rather than specific entities.

We calculate a score for each label within a class by passing the model's predictions through a softmax layer. Softmax converts the predictions into probabilities, assigning a likelihood that a given headline belongs to each label. We use these probabilities as the scores for the labels.

4.4.2 News Analysis

We analyze the news using fine-tuned RoBERTa model due to its superior performance compared to other models. We compare the war and peace journalism frames across news publishers for different entity groups. We use the four entity groups discussed in Section III-B, namely, PALESTINE, HAMAS, ISRAEL, and OTHERS. To improve robustness, we discard headlines with overlapping entity groups, i.e., if the same headline mentions entities from two different groups, we discard it.

We evaluate the victim and villain portrayal of each entity group and identify differences between news publishers. For this, we use all the news headlines and their corresponding scores for each entity group using the model trained to identify the victim and the villain. We compare the mean scores across news publishers. Additionally, we compare the timelines for villain and victim portrayal of each entity by different publishers. We use a Gaussian filter to smooth the data timeline for an effective comparison. We use SciPy[1] to apply the Gaussian filtering.

We conduct statistical tests to assess if the differences between news publishers covering the same entity are significant. For this, we drop the score for None from each frame and normalize the scores between war and peace frames. To choose the appropriate statistical test, we conduct a normality test using the Shapiro-Wilks normality test (Shapiro and Wilk 1965). Since the distributions are not normal, we choose the Mann-Whitney U test (Mann and Whitney 1947). The Mann-Whitney U test compares two independent groups without assuming a normal distribution, making it suitable for non-parametric data. Moreover, it doesn't require paired data.

To quantify the magnitude of the differences between compared distributions, we compute the effect size for each test. We use Rank Biserial Correlation (RBC) effect size (Cureton 1956), suitable for non-parametric statistical tests. We prefer RBC because of its robustness for ranked statistical tests such as the Mann-Whitney U test. The RBC effect size ranges between -1 and 1, with 0–0.3 indicating a small effect, 0.3–0.5 indicating a medium effect, and 0.5–1.0 indicating a large effect. Negative and positive values indicate the direction of the relationship, i.e., whether the first group tends to have higher ranks than the second group (positive) or the second group

[1]https://docs.scipy.org/doc/scipy

tends to have higher ranks (negative).

The codebase and dataset for this work is publicly available on Github[2]

Entity	News Source		War and Peace Frames								Role			
	Pub 1	Pub 2	Emotiveness		Effect		Source		Context		Victim		Villain	
			p-val	ES	*p*-val	ES	*p*-val	ES	*p*-val	ES	*p*-val	ES	*p*-val	ES
Palestine	Alj	BBC	**1.92e-06**	-0.272	**9.36e-03**	0.149	**5.39e-04**	-0.198	**3.96e-02**	0.118	1.30e-01	0.087	6.94e-01	0.023
	BBC	Fox	**4.52e-02**	0.165	**1.02e-06**	-0.403	**9.96e-02**	-0.136	**1.16e-13**	-0.611	**3.18e-08**	0.456	**1.94e-05**	-0.352
	Alj	Fox	1.05e-01	-0.118	**1.61e-04**	-0.275	**3.22e-06**	-0.339	**5.97e-15**	-0.568	**1.69e-14**	0.559	**2.39e-07**	-0.376
Hamas	Alj	BBC	**4.90e-02**	-0.148	**1.81e-02**	0.178	9.71e-01	-0.003	**8.25e-03**	0.199	4.28e-01	-0.060	**9.02e-05**	-0.295
	BBC	Fox	**1.53e-12**	0.444	**5.68e-08**	-0.340	2.50e-01	-0.072	**3.34e-09**	-0.371	8.17e-01	0.015	**6.50e-08**	-0.339
	Alj	Fox	**1.08e-07**	0.340	**2.73e-03**	-0.192	**5.01e-02**	-0.125	**5.33e-04**	-0.222	5.26e-01	-0.041	**8.49e-22**	-0.614
Israel	Alj	BBC	**7.95e-16**	-0.232	**1.13e-03**	-0.094	**7.74e-14**	-0.215	**1.60e-04**	-0.109	**1.43e-19**	-0.260	**2.22e-50**	0.430
	BBC	Fox	**6.22e-42**	0.423	**9.50e-18**	-0.267	6.81e-01	-0.013	**4.02e-33**	-0.374	**3.19e-17**	-0.263	**1.68e-05**	0.134
	Alj	Fox	**2.80e-10**	0.161	**8.46e-45**	-0.359	**7.22e-23**	-0.252	**4.48e-72**	-0.459	**1.24e-94**	-0.528	**6.77e-115**	0.583
Others	Alj	BBC	**1.77Ee08**	-0.186	**8.15e-05**	0.130	**9.14e-03**	-0.086	**4.07e-04**	0.117	**8.89e-04**	-0.110	**1.96e-03**	-0.102
	BBC	Fox	**5.29e-08**	0.251	**1.07e-05**	-0.203	**1.33e-09**	-0.280	**1.69e-16**	-0.380	**8.91e-03**	-0.121	2.04e-01	-0.059
	Alj	Fox	3.77e-01	0.040	**4.27e-02**	-0.092	**9.55e-17**	-0.375	**4.09e-11**	-0.298	**4.36e-08**	-0.247	**4.91e-04**	-0.157

Table 4.5: Mann Whitney U test p-values (*p*-val) and effect size (ES) between news publishers. p-values < 0.01 are **bolded**.

4.5 Findings

4.5.1 Model Evaluation

Table 4.4 presents the results for different fine-tuned models on our dataset. RoBERTa outperforms most models, achieving the highest F1 scores for all war and peace frames, except for the Emotiveness frame, where ConfliBERT (84.83) and BART (84.18) slightly outperform RoBERTa (83.83). For role labeling (victim and villain), RoBERTa performs best in identifying the victim, while ConfliBERT surpasses RoBERTa in identifying the villain by a small margin (81.74 for ConfliBERT vs. 81.16 for RoBERTa). Although ConfliBERT matches RoBERTa for most classes, it shows a sharp decline in effect frames (61.46 for ConfliBERT vs. 71.55 for RoBERTa). Based on these results, we selected RoBERTa to conduct news analysis on the larger dataset, as it consistently performs well and shows fewer signs of overfitting.

[2] https://github.com/ahaque2/War_and_Peace_Journalism

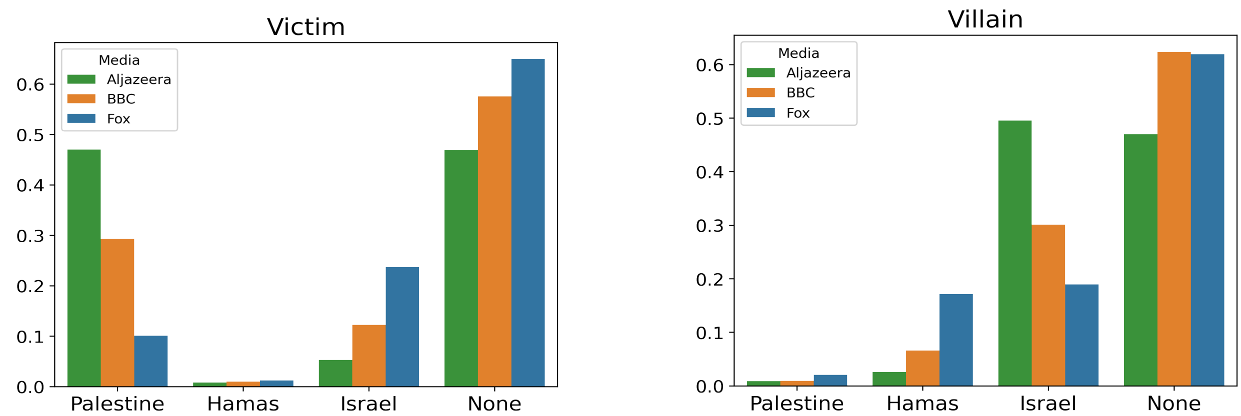


Figure 4.2: Mean victim and villain scores in news headlines.

4.5.2 Victim and Villain

Figure 4.2 compares how different news publishers portray victims and villains. The None category has the highest scores for both victim and villain, indicating that most news reports do not explicitly depict an entity as either a victim or a villain. However, when entities are framed this way, distinct patterns emerge across publishers. Fox News tends to be supportive of Israel, portraying Israel as a victim and Hamas as a villain. Although Fox News sometimes depicts Israel as a villain, it does so less frequently than Aljazeera and BBC. In contrast, Aljazeera portrays Palestine as a victim and Israel as the villain. BBC presents a more balanced portrayal of all entities, with its victim and villain scores falling between those of Aljazeera and Fox News.

All news publishers show low scores in portraying Palestine as a villain and Hamas as a victim. However, Fox News assigns a higher villain score to Palestine and a higher victim score to Israel compared to Aljazeera, reflecting differences that align with each publisher's editorial stance. Despite these variations, the differences are smaller than for other entities and roles, suggesting some consensus on how these particular entities are framed in these roles.

Table 4.5 presents the results of statistical significance tests comparing news coverage across different publishers. Most comparisons yield a low p-value (<0.01), indicating significant differences in coverage, with effects ranging from small to large. The largest effect is observed in the portrayal of Hamas as a villain, where Fox News (−0.614) depicts Hamas as a villain significantly more than Aljazeera, whereas differences between BBC–Fox and Aljazeera–BBC indicate moderate effects. In contrast to its villain portrayal, the victim portrayal of Hamas shows much less variation across publishers, with all comparisons yielding a small effect.

Palestine shows differences with higher effects for its victim portrayal than villain, with large effects for victim portrayal between Aljazeera and Fox, and low to moderate effects for other comparisons. For Israel, both victim and villain portrayals show differences with large

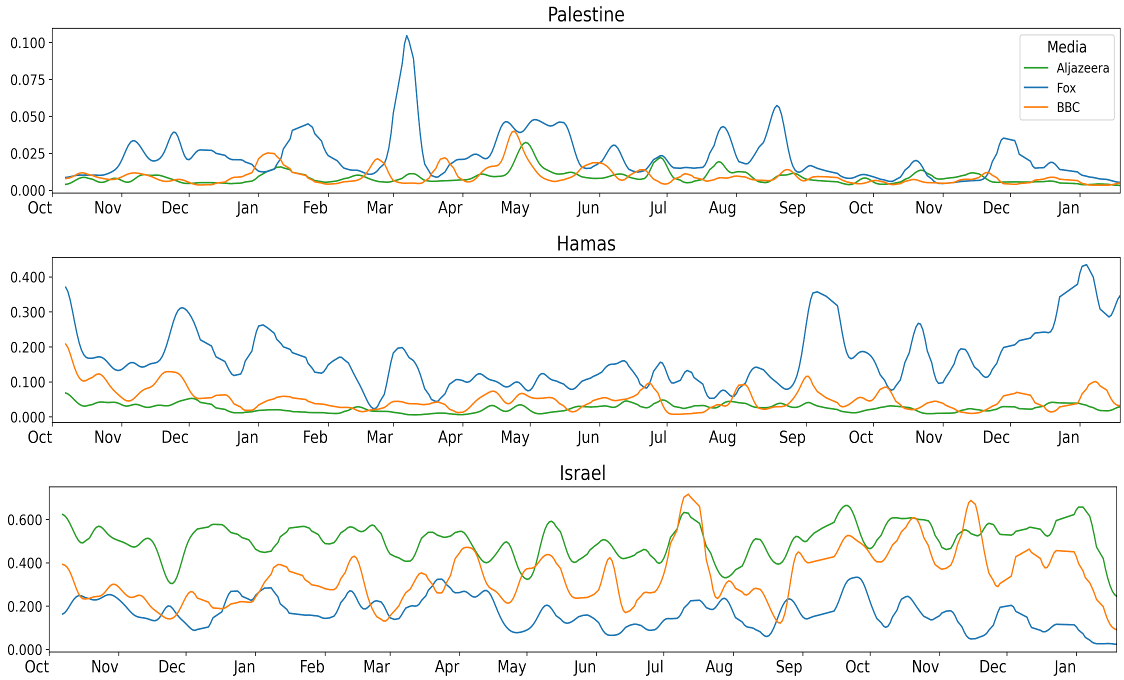


Figure 4.3: Daily mean villain scores.

Figure 4.4: Daily mean victim scores.

effects between Aljazeera and Fox. In contrast, coverage of Others does not exhibit large effects, suggesting more consistent reporting across publishers for non-conflicted entities.

Figure 4.3 illustrates how different news publishers have portrayed villains over time. The portrayal of Palestine as a villain follows a similar timeline across all three publishers, aligning with observations from Figure 4.2. Although BBC and Aljazeera's coverage patterns are more alike, Fox occasionally exhibits peaks where Palestine is depicted as a villain. Hamas is consistently framed as a villain more often by Fox than by BBC or Aljazeera. However, Fox's villain score for Hamas is higher at the beginning and end of the timeline, with a dip in the middle, indicating shifts in reporting over time. Aljazeera consistently portrays Israel as a villain more consistently than Fox and BBC. Interestingly, BBC's coverage initially aligns more closely with Fox but later shifts toward Aljazeera's. This suggests that BBC's coverage of Israel was initially more sympathetic but became less so in the later stages of the conflict.

Figure 4.4 illustrates the portrayal of the victim over time across three publishers. Hamas is portrayed similarly by all three, with the lowest victim portrayal scores and occasional peaks, which are consistent across the publishers. Coverage of Palestine and Israel diverges, with Aljazeera consistently portraying Palestine as a victim, whereas Fox presents Israel as the victim. BBC maintains a relatively balanced stance, portraying Palestine as a victim more than Fox but less than Aljazeera, and Israel as a victim more than Aljazeera but less than Fox.

4.5.3 War and Peace Frames

Figure 4.5 compares the war and peace frames for different entities across the three news publishers.

For the emotiveness frame, all three news publishers present more objective headlines than sensationalized ones. However, we observe differences across entities. Fox presents more sensationalized headlines for Hamas and Israel than Aljazeera or BBC. In contrast, Aljazeera presents more sensationalized headlines for Palestine than Fox. BBC consistently presents less sensationalized headlines than both Fox and Aljazeera for all entities.

BBC reports visible effects more frequently than Fox and Aljazeera for all entities, except for Israel, where Aljazeera reports more visible effects. Fox reports more invisible effects than both BBC and Aljazeera, particularly for Palestine, and reports more invisible effects overall than visible ones.

For the source frame, Fox reports more stories from elite sources than people's stories, except for Palestine, where it is proportional. Fox's use of elite sources is greater than that of both BBC and Aljazeera. In contrast, Aljazeera reports more people's stories for Palestine and proportionally for other entities. BBC provides relatively balanced reporting, using a mix of

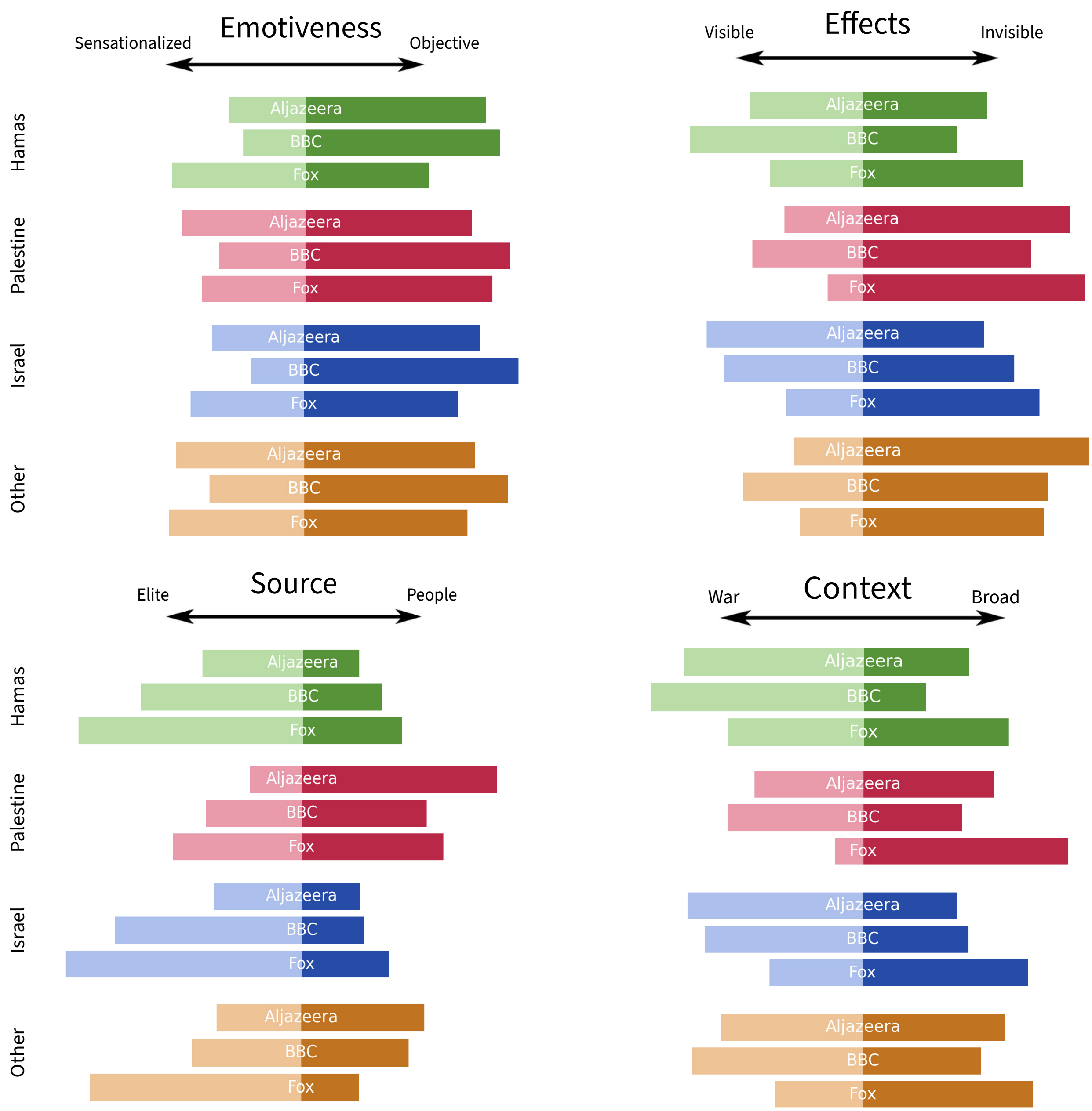


Figure 4.5: War and Peace Frames.

different source frames for different entities.

For the context frame, Fox reports the most headlines that provide a broader context, especially for Palestine. BBC provides the most war-related context, except for Israel, where Aljazeera reports more war context.

Table 4.5 presents the results for statistical significance tests comparing war and peace frames across news publishers. Most comparisons yield a low p-value, indicating significant differences across publishers. The effects range from small to large, with some of the smallest effects observed for Others and the largest for Palestine.

4.6 Discussion

Our analysis of news coverage on the ongoing war in Gaza reveals key differences across publishers and provides valuable insights into conflict reporting. We find significant differences in the portrayal of victims and villains, with different news publishers portraying different entities as victims and villains consistent with previous research (Heni and Chandra 2022; Doufesh and Briel 2021), suggesting a partisan coverage. Although the portrayal of victims and villains shows signs of partisan coverage, the use of war and peace frames presents a less conclusive picture, as all publishers use both frames.

Shift in BBC's coverage. The timeline of villain and victim portrayals reveals notable shifts over time. Initially, BBC's coverage of Israel as a villain aligns more with Fox than Aljazeera. However, as the conflict progresses, BBC's stance shifts, increasingly portraying Israel as a villain. This change coincides with key external events that drew widespread criticism of Israel. BBC's coverage started shifting around late March following Israel's attack on Al Shifa hospital, which was widely condemned. The increase in Israel's villain portrayal in April and May aligns with Israel's attack on Rafah, a designated safe zone sheltering many Palestinians. The peak in July corresponds with Israel's invasion of Khan Younis and Rafah, which the UN described as the largest displacement order since the war began. Another peak in late November coincides with the International Criminal Court issuing arrest warrants for Netanyahu and his cabinet members for alleged war crimes.

Fox's use of peace frames. The use of war and peace frames varies across publishers but presents a mixed picture. All publishers incorporate both war and peace frames, but the choice of using a frame often depends on factors beyond a deliberate intent to promote war or peace. For example, Fox reports more on invisible effects and broader context (peace frames) than Aljazeera and BBC but relies more heavily on elite sources (war frame). Although the use of a peace frame may seem to suggest a framing supporting of peace, its not always the case. Many of Fox's reports on invisible effects and broader context focus on U.S. reactions to the Gaza war, such as pro-Palestine protests on college campuses and rising antisemitism. While it acknowledges growing pro-Palestine demonstrations, its coverage remains largely supportive of Israel, often framing Palestinian supporters in a negative light. Some examples of Fox headlines describing these protests are as follows.

Anti-Israel agitators shut down traffic, disrupt cities all across US in demand for Gaza cease-fire.

'Radical' pro-Palestinian groups increasingly target houses of worship for protests in alarming trend.

Fox's use of war frame. Fox relies on elite sources more than both BBC and Aljazeera and more than people's stories, except in coverage of Palestine, where the two are proportionally balanced. Studies examining media coverage of conflicts have identified reliance on elite sources as a key indicator of war journalism (Lee and Maslog 2006; Workneh 2011). Moreover, public opinion is heavily shaped by exposure to elite discourse, especially on political matters (Zaller 1992b). This suggests that Fox's preference for elite sources could signal a war-framing approach in its reporting on this conflict. However, Tenenboim-Weinblatt et al. (2016) caution against concluding a war framing based solely on the presence of elite sources, emphasizing the need for additional context to make such an inference. They argue that the use of elite sources may not be sufficient to conclude a war framing, as some elite sources promote peace. Moreover, even elite sources have a gradation that war and peace journalism overlooks. Additionally, it matters how the elite source is used. For instance, a mention in the headline has far more influence than if the elite source is mentioned at the end of the news article (perhaps to provide a balance).

Limitations of war and peace journalism framework. Fox's use of peace frames while still promoting a war-supporting narrative highlights a key limitation of the war and peace journalism framework. The choice of war or peace frames is not solely a deliberate editorial decision but is also influenced by factors such as a story's perceived newsworthiness. A news publisher within the war zone is more likely to report more on visible effects due to its immediate newsworthiness to its audience. In contrast, a publisher geographically distant may focus more on domestic reactions and the war's broader influence. This difference in coverage priorities may lead distant publishers to use more peace frames, such as highlighting invisible effects or providing broader context. For example, many of Fox's peace frames, such as invisible effects and broader context, appear to be driven more by the prominence of related events than by an intention to promote peace. As a US-based news publisher, Fox often publishes news about domestic reactions to the war, which often fall under broader contexts or invisible effects. However, the focus on invisible effects and broader context doesn't guarantee a peace framing, underscoring an important limitation of the war and peace journalism framework.

4.6.1 Threats to Validity

Although we implement consistent coding practices and perform thorough analyses, potential threats to validity remain.

First, the interrater agreement for some classes isn't substantial (i.e., Fleiss Kappa < 0.60). While this may seem like a threat to the validity of our dataset, we argue this demonstrates the challenging nature of the task and emulates how news is perceived in the real world. The low annotator agreement reflects the subjective nature of applying the framework to news headlines, as framing interpretations can vary widely. Since real-world news consumers also have diverse perspectives and little formal training, crowdsourcing annotations for this task create conditions that closely mirror how news headlines are understood in practice. Rather than enforcing a consensus through rigorous training, this variation captures the complexity of framing in journalism.

To ensure robustness in our annotations, we conduct annotations in three phases. Phase one involves annotating 100 headlines among a small group, followed by discussions intended to improve the annotation instructions. Phase two involves training the crowdsource annotators by providing 50 headlines to annotate based on the annotation instructions from phase one. Additionally, we ask the annotators to explain their annotations, which we use to improve the instructions further. Phase three, where the crowdsource annotators annotate the whole dataset. To ensure robustness in our annotations, we discard all annotations from phases one and two. Additionally, we use majority vote for aggregation and discard headlines where no consensus is achieved.

Second, we annotate only headlines and not the entire news. While the full article may present a different perspective or framing than the headline, the choice of only focusing on the headline is deliberate. This is because headlines play a key role in shaping how readers interpret the news and, in the age of social media, are often used as clickbait, making their framing especially impactful. Additionally, most prior empirical studies on war and peace journalism have focused on full articles, with little attention to headlines, creating a research gap that this work intends to address.

CHAPTER

5

UNDERSTANDING DYNAMICS OF POLARIZATION VIA MULTIAGENT SOCIAL SIMULATION

5.1 Introduction

Polarization is a societal problem since it makes rational decision-making and resource allocation difficult. Not only does it pose objective challenges to society and business, it also causes widespread attitudes and behaviors that are antivax, antimask, and antiscience. This is worsened by the internet, which enables fast information diffusion across traditional boundaries. Specifically, social media influences users in subtle ways, especially regarding politics (Nahon 2015). Moreover, research has shown a strong correlation between online and offline political participation (Johnson et al. 2020; Bode et al. 2014) which further raises concerns.

To analyze polarization in social networks and understand its dynamics based on information sharing, we design and implement a multiagent social simulation. We focus on two factors identified by prior research that influence polarization. *First*, selective exposure to congenial (attitude-conforming) information exacerbates confirmation bias, polarizing opinions further (Stroud 2010; Garrett et al. 2014; Kim 2015; Westerwick et al. 2017). Selective exposure arises in and strengthens echo chambers, wherein a person encounters only beliefs or opinions that coincide with their own so that their existing views are reinforced and alternative ideas are suppressed. Conversely, cross-cutting exposure (to uncongenial i.e., attitude-disconfirming information) has a depolarizing effect (Kim 2015), though with caveats (Garrett et al. 2014; Kim 2019).

Second, user tolerance for ideas that contradict existing beliefs mitigates polarization (Coscia and Rossi 2022).

We analyze the effects of selective exposure and tolerant users on polarization among users at large. Specifically, we investigate the following research questions.

$\text{RQ}_{\text{tolerance}}$. *Does higher tolerance among users in a social network help mitigate polarization?*

$\text{RQ}_{\text{exposure}}$. *Does selective exposure to congenial information contribute to polarization?*

We develop a multiagent social simulation to investigate these research questions.

To address $\text{RQ}_{\text{tolerance}}$, we model tolerant users by having a higher tolerance level toward opposing views. We operationalize tolerance in users using social judgment theory (Sherif and Hovland 1961), which defines tolerant people as those having a wider latitude of noncommitment. For $\text{RQ}_{\text{exposure}}$, we emulate selective exposure by filtering posts based on the receiving user's stance toward the post.

We observe higher user tolerance mitigates polarization but achieves lower user satisfaction than users with lower tolerance. Surprisingly, higher tolerance leads to a more homophilic social network. Higher selective exposure leads to more polarization, a more homophilic social network, and a higher aggregate user satisfaction but fewer satisfied users.

Analyzing polarization dynamics based on information sharing on social media can help us identify potential interventions. Since most algorithmic content filtering in use today selectively exposes users to congenial content, this work can help us better understand the long-term effects of varying levels of such exposure. Our findings suggest avenues for further theoretical development in tandem with the consideration of interventions to reduce polarization in online social networks.

Organization. The rest of the paper is organized as follows: Section 5.2 describes the background and discusses related work. Section 5.3 explains our methodology, including definitions and the simulation design, assumptions and limitations. Section 5.4 details the experimental setup, results of our experimentation, and statistical analysis of the results. Section 5.5 discusses the findings Section 5.6 presents the threats to the validity of this work. Section 5.7 concludes with future directions.

5.2 Background and Related Work

The Theory of Cognitive Dissonance (Festinger 1957) asserts that when a person is confronted with contrasting ideas, it causes psychological discomfort making that person more selective in their information consumption, potentially causing confirmation bias. *Confirmation bias* is the tendency of people to accept "confirming" evidence at face value while subjecting "disconfirming" evidence to critical evaluation (Lord et al. 1979), resulting in people gravitating toward information that aligns with (confirms) their existing views. Bias exists in the selection and sharing of information, especially news (Hart et al. 2009; Knobloch-Westerwick 2014).

Selective exposure is a tendency of people to choose and spend more time on information that is consistent with their existing beliefs (Klapper 1960; Redlawsk 2002; Taber and Lodge 2006), though some prior works suggest that partisan selective exposure may be a myth (Kinder and Sears 1981; Zaller 1992b). Freedman and Sears (1965) argue against voluntary selective exposure in favor of *de facto selectivity*. They claim that most examples of selectivity in mass communication can be attributed to complex factors such as demography, education, social connections, and occupation, which are incidental to their supportiveness to the receiver's existing beliefs. People prefer supportive information in some situations while dissonant information in other situations (Hargittai et al. 2008). Individuals with strong preferences are more likely to spend more time reading negative (uncongenial) information about their choice (Meffert et al. 2006), perhaps to critique it (Hargittai et al. 2008).

5.2.1 Social Media and Politics

The number of users on social media platforms has increased rapidly over the years. Only 8% of the Internet users in the US used some social networking platform in 2005 (Lenhart 2009), whereas in 2021, 69% use Facebook, and 40% use Instagram (Auxier and Anderson 2021). The use of social networking sites for political discussions has also increased over the years. Social media is now among the most common ways in which people, particularly young adults, obtain their political news (Infield 2020). A meta-analysis from 36 past studies assessing the relationship between social media use and participation in civic and political life found a positive correlation between the two, with more than 80% of the coefficients as positive (Boulianne 2015). Polarization measured based on online social interactions shows a good correlation with offline polarization (Morales et al. 2015). Adults who use social networking platforms as a political tool are more likely to participate in politics (Bode et al. 2014). This is true across various cultural and geographical boundaries, including empirical evidence from the US (Infield 2020), Pakistan (Ahmad et al. 2019), and Taiwan (Zhong et al. 2022)

Selective exposure to political information is correlated with polarizing people's opinions to align with the values of the political party they support (Stroud 2010; Garrett et al. 2014; Kim 2015; Westerwick et al. 2017). Though the causal direction, i.e., whether selective exposure leads to polarization or the other way around, is less obvious (Stroud 2010). Stroud (2010) investigate the causal relationship between partisan selective exposure and polarization and find strong evidence suggesting selective exposure leads to polarization while finding limited evidence suggesting the reverse causal direction. Schkade et al. (2007) find that intragroup deliberation on social issues among like-minded people leads to more extreme and less diverse ideological beliefs, whereas Bail et al. (2018) observe that exposure to opposing views on social media can increase political polarization. Habitual online news users are less likely to exercise selectivity to get attitude-consistent exposure, which reduces their likelihood of participating in the political system (Knobloch-Westerwick and Johnson 2014). The longer individuals spend on attitude-consistent content associated with biased sources, the more immediate attitude reinforcement occurs, and its influence can be detected even after a couple of days of exposure (Westerwick et al. 2017).

Cross-cutting exposure refers to being exposed to oppositional viewpoints. Cross-cutting exposure in social networks fosters political tolerance and makes individuals aware of legitimate rationales for oppositional viewpoints (Mutz 2002b). Exposure to disagreeing viewpoints contributes to people's ability to generate reasons, particularly why others might disagree with their view (Price et al. 2002). Kim and Chen (2016) find that exposure to cross-cutting perspectives results in a higher level of political engagement, though this increase may depend on the social media platform used.

Cross-cutting exposure, widely assumed to encourage an open and tolerant society, is not necessarily the environment that produces enthusiastically participatory individuals. People belonging to social networks involving greater political disagreement are less likely to participate in politics (Mutz 2002b,a). Constant exposure to disagreement may necessitate trade-offs in other social network characteristics such as relationship intimacy and frequency of communication (Mutz 2002b). Conflict-avoiding individuals, in particular, are more likely to respond negatively to cross-cutting exposure by limiting their political participation to avoid confrontation and putting their social relationships at risk (Mutz 2002a).

Garrett et al. (2014) examine survey data following elections in the US and Israel and find consistent results despite cultural differences. Their findings suggest that pro and counter-attitudinal information exposure has a distinct influence on perceptions of and attitudes toward members of opposing political parties.

Mutz (2002a) analyzes the consequences of cross-cutting exposure on political participation. They find that people whose social networks involve greater political disagreement are less likely to participate in politics and are more likely to hold politically ambivalent views.

Though many studies have investigated polarization using empirical data from social media, a common limitation has been that past studies either look at one-time exposure or study these effects in isolation. For instance, Stroud (2007) studies the effects of selective exposure using empirical evidence but relies on data from one-time exposure and studies the immediate effects without differentiating the long-term effects. However, the evidence from past studies suggests that political participation and its effect is a long-term process that unfolds over time based on multiple exposures (Gerber et al. 2003; Valentino and Sears 1998). Further, existing research has focused on the effect at an individual level, i.e., relying on self-reported data on how an individual's stance is influenced by exposure to potentially polarizing content. However, self-reporting is susceptible to user bias and overlooks how changes in one part of the social network can influence other parts.

5.2.2 Multiagent Social Simulation

Many earlier models on opinion and influence propagation are based on a centralized diffusion process, overlooking the decentralized nature of information diffusion in social networks.

Kempe et al. (2003) design two fundamental diffusion models for influence maximization, namely, the Independent Cascade Model (ICM) and the Linear Threshold Model (LTM). Influence in these models is transferred through the correlation graph starting from a set of seed nodes (activated nodes). Influence decreases when hopping further away from the activated node.

Jiang et al. (2017) design a preference-aware and trust-based influence maximization model called the Preference-based Trust Independent Cascade Model (PTICM) that takes into account user preferences and trust between users in computing influence propagation.

Li et al. (2019) design a novel agent-based seeding algorithm for influence maximization named Enhanced Evolution-Based Backward selection that models individual user preferences and social context based on social influence and homophily. Their results suggest that individuals are influenced by their social context much more than retaining their own opinions. Though the Prior Commitment Level (PCL) of a user is an essential factor for influence propagation, users tend to revise their PCL over time.

Chen et al. (2020a) propose a group polarization model based on the SIRS epidemic model and factor in the relationship strength based on the J-A (Jager and Amblard) model. They use a BA network model due to its closeness to the real-world social network structure and a Monte Carlo method to conduct simulation experiments.

Kozitsin and Chkhartishvili (2020) develop an agent-based model to explore how agents' activity patterns affect the formation of echo chambers. They use a personalizing system algorithm to control mutual interactions among agents and decide what information the agents are exposed to. They find that the critical parameter that guides agents' opinion dynamics is the probability of publishing a post, i.e., agents who often publish posts tend to enter echo chambers.

Hązła et al. (2019) use a geometric model of polarization and demonstrate that societal opinion polarization often arises as an unintended byproduct of influencers attempting to promote a product or an idea. Gaitonde et al. (2021) extend this work to show that the exact form of polarization in such models is quite nuanced. Even when strong polarization does not hold, weaker notions of polarization can attain nonetheless.

Baumann et al. (2020) propose a radicalization model that uses a reinforcement mechanism to drive opinions to extremes starting from moderate initial conditions. They show that the transition from a global consensus to a radicalized state is mostly governed by social influence and the controversialness of topics discussed.

Wang et al. (2019) model a rumor-propagation framework based on information entropy to understand information distortion and its polarization effects in social networks. They find that mass polarization toward a positive or negative consensus occurs when a synergistic mechanism between preferential trust and polarization tendencies is sustained. The segregation of the population into groups of different polarities happens under certain conditions.

We design a multiagent social simulation to emulate information diffusion on social networks. We model user behavior based on existing social science theories and empirical evidence from prior studies.

5.3 Methodology

We now describe our social simulation model and agents' interaction.

5.3.1 Social Simulation Definitions

Definition 1 (Social Network) *A* social network *is an undirected graph with nodes representing users and the links connecting the nodes representing a relationship between two users.*

A social network is represented as G = *(nodes, edges)*, where *nodes* = $\{a_1, \ldots, a_n\}$ are users and *edges* = $\{(a_1, a_2), (a_4, a_9), \ldots, (a_x, a_y)\}$ represent a direct connection between pair of users in the social network. An agent can only interact with its neighbors in the social network.

Definition 2 (Agent) *An* agent *represents a user in the social network.*

Each agent is independent and has attributes defining its preferences such as *user activity*, and *sharing preference*. User activity captures how active an agent is, and sharing preference captures agents' willingness to share a post on the social network. Both range over [0, 1] (0 represents most inactive/unwilling and 1 most active/willing). An agent can take two actions: sharing a post and providing sanctions to posts.

Definition 3 (Post) *A* post *is a message shared by an agent with its neighbors in the social network.*

Agents in a social network interact by sharing posts that can be represented as *Post = (a, t, s)*, where *a* is the author, *t* is the topic mentioned in (or discussed in) the post, and *s* is the stance of the post towards the topic (continuous value in [−1, 1], where −1 represents extreme opposition and 1 extreme support for the issue).

A post serves as a timestep and is used to track changes in the social network over time. Updates to the social network and agent's attributes are made after a post has completed diffusion in the social network (i.e., it has reached as many agents as possible).

Definition 4 (Sanction) *A* sanction *is an agent's reaction to a post it receives.*

Sanctioning provides a foundation for how participants in a sociotechnical system (STS) may seek to influence each other's decision-making and steer the STS toward their preferred direction (Nardin et al. 2016). Agents provide positive sanctions to congenial posts and negative to uncongenial posts based on their stance on a given topic being discussed in the post. Sanctioning is analogous to providing likes and comments to a post and captures whether a user approves (likes) or disapproves (dislikes) the topic in a received post.

Definition 5 (Issue) *An* issue *refers to the topic being discussed in a post.*

Issues are predefined, and all agents hold a stance on each issue. An agent's stance toward an issue is represented as a continuous value between [−1,1], with −1 indicating extreme opposition and 1 extreme support for the issue. Each agent has an overall POV (point-of-view) that depends on its stance on various issues. The POV of an agent is computed as the mean of its stance on all issues. POV ranges between [−1, 1], with −1 representing extreme support for POV-1 (<0), 0 means neutral POV, and 1 extreme support for POV-2 (>0).

With respect to a post, an agent can be in one of the four states:

1. *Nonreceiver:* Agents who have not yet received the post (all agents other than the author are in this state at the start of the simulation);

2. *Receiver:* Agents who have received the post (but not yet shared it);

3. *Spreader:* Agents who have shared the post with their friends; and

4. *Disinterested:* Agents who received the post but chose not to share it further and lost interest in the post.

5.3.2 Social Simulation Model

The simulation starts with an agent (a_x) sharing a post (p_k) with its neighbors in the social network. The receiver then decides whether to share the received post further with a probability of sharing that depends on the content of the post and the receiver's preferences. An agent's preference involves its sharing preference, how active the agent is on the social network, and the agent's stance toward the issue (supporting vs. opposing). The content of a post includes the issue mentioned in the post and the post's stance toward the issue. Equation 5.1 describes the computation for sharing probability sP(a_x, p_k) for the agent a_x to share the post p_k it received.

$$\mathrm{sP}(a_x, p_k) = c_1 \times \mathrm{uA}(a_x, p_{k-1}) \times |\mathrm{uS}(a_x, i, p_{k-1}) \times \mathrm{pS}(p_k, i)| \times \mathrm{sPref}(a_x, p_{k-1}) \tag{5.1}$$

c_1 is a constant, a_x is the receiver, p_k is the k^{th} post being shared in the social network, and i is the issue being discussed in the shared post. $\mathrm{uA}(a_x, p_{k-1})$ is the user activity of user a_x before the post p_k is shared, $\mathrm{uS}(a_x, i, p_{k-1})$ is the user a_x's stance towards issue i before the post p_k is shared, $\mathrm{pS}(p_k, i)$ is the stance of the post towards issue i, and $\mathrm{sPref}(a_x, p_{k-1})$ is the sharing preference of user a_x before the post p_k is shared. An agent with low $\mathrm{sPref}(a_x, p_{k-1})$ is more likely not to share a post further and may enter the state *Disinterested*. Disinterested agents are not candidates for sharing the post (p_k) further.

The agents who receive the post provide a sanction. Sanctions can be positive or negative. Sanctions by the receiver depend on how active the receiver is, its stance toward the issue at hand, and the post's stance toward the issue. Sanction by an agent a_y for a post p_k it received from agent a_x is computed using Equation 5.2.

$$\mathrm{Sanc}(a_y, p_k, a_x) = c_1 \times \mathrm{uA}(a_y, p_{k-1}) \times \mathrm{uS}(a_y, i, p_{k-1}) \times \mathrm{pS}(p_y, i) \tag{5.2}$$

$\mathrm{Sanc}(a_y, p_k, a_x)$ is a sanction provided by agent a_y for the post p_k it received from agent a_x. Sanction scores affect user activity and the stance of each agent towards an issue. Agents prefer positive sanctions (social acceptance), which increases their activity on the platform, while negative sanctions discourage agents from sharing their views in the future, hence reducing their participation (user activity). The update in user activity depends on the sanctions received by an agent for the posts it shared. An agent's user activity (uA(a_x, p_k)) after sharing a post p_k is computed using Equation 5.3.

$$\mathrm{uA}(a_x, p_k) = \mathrm{uA}(a_x, p_{k-1}) + c_2 \times \sum_{a_i \in \mathrm{neighbor}(G, a_x, p_k)} \mathrm{Sanc}(a_i, p_k, a_x) \tag{5.3}$$

c_2 is a constant, uA(a_x, p_{k-1}) represents the user activity of agent a_x before the post p_k is shared, and uA(a_x, p_k) represents the user activity of agent a_x after the post p_k is shared. neighbor(G, a_x, p_k) refers to all neighbors of agent a_x in the social network G that receive the post p_k directly from agent a_x.

An agent's stance toward an issue is influenced by the sanctions it receives from other agents. We model this shift in the stance of an agent using social judgment theory (SJT) (Sherif and Hovland 1961), which describes how individuals change their position when confronted with a competing position on a given issue. According to SJT, an individual shifts their stance in the direction of the competing stance if the competing stance falls within their *latitude of acceptance* (assimilation). In contrast, they will shift away from the competing stance if the competing stance falls beyond their *latitude of rejection* (contrast). For instance, for an agent a_x, that has a stance of uS(a_x, i, p_k) towards issue i, a threshold determining the latitude of acceptance u_{xi} and a threshold determining the latitude of rejection t_{xi} with $t_{xi} > u_{xi}$. When this agent a_x interacts with another agent a_y, the following rules are applied to compute the shift in the stance of agent a_x towards an issue i.

$$\mathrm{diff\_Stance}(a_x, a_y, i, p_k) = |\mathrm{uS}(a_x, i, p_k) - \mathrm{uS}(a_y, i, p_k)| \tag{5.4}$$

diff_Stance(a_x, a_y, i, p_k) is the absolute difference in the stances of agent a_x and agent a_y on issue i as the post p_k is being shared.

$$\begin{aligned}
&\text{If } \mathrm{diff_Stance}(a_x, a_y, i, p_k) < u_{xi}, && \delta\mathrm{uS}(a_x, a_y, i, p_k) = \mu \times (\mathrm{uS}(a_y, i, p_k) - \mathrm{uS}(a_x, i, p_k)) \\
&\text{If } \mathrm{diff_Stance}(a_x, a_y, i, p_k) > t_{xi}, && \delta\mathrm{uS}(a_x, a_y, i, p_k) = \mu \times (\mathrm{uS}(a_x, i, p_k) - \mathrm{uS}(a_y, i, p_k)) \\
&&& \text{else } \delta\mathrm{uS}(a_x, a_y, i, p_k) = 0
\end{aligned} \tag{5.5}$$

μ represents the strength of the influence between two agents. We assume the same strength of influence between all pairs of connected agents in the social network; hence the value of μ is 1. The shift in the stance of an agent a_x for sharing posts p_k on issue i is computed using the received sanction scores and the difference in stance (toward the issue at hand) between the author or spreader (i.e., a_x) of the post, and the receiver (i.e., a_y) (Equation 5.6).

$$\Delta S(a_x, a_y, i, p_k) = c_2 \times \frac{\mathrm{Sanc}(a_y, p_k, a_x)}{\delta \mathrm{uS}(a_x, a_y, i, p_k) + 1} \tag{5.6}$$

$\Delta S(a_x, a_y, i, p_k)$ is the shift in stance (of agent a_x) due to a sanction (by agent a_y) for a post p_k it shared on issue i.

User stance after sharing post p_k can be computed using Equation 5.7.

$$\mathrm{uS}(a_x, i, p_k) = \mathrm{uS}(a_x, i, p_{k-1}) + \sum_{a_j \in \mathrm{neighbor}(G, a_x, p_k)} \Delta S(a_x, a_j, i, p_k) \tag{5.7}$$

$\mathrm{uS}(a_x, i, p_{k-1})$ is the stance of the agent a_x on issue i before it shares post p_k, and $\mathrm{uS}(a_x, i, p_k)$ is the stance of an agent a_x on issue i after the posts p_k is shared and sanctions for it received from all other agents. The maximum allowed change in stance due to one post is 0.20, and we bound user stance within $[-1, 1]$ by restricting the values.

The codebase[1] of our social simulation is publicly available.

5.3.3 Agent Goals and Actions

The simulation progresses with agents sharing posts with other agents, causing each post to diffuse further in the social network. Each post receives a sanction from all agents that receive it, and these sanctions, in turn, influence its authors' (spreaders') activity score and stance toward various issues. An agent supports a POV with which its aggregate stance toward various issues is in agreement. Agents can take two actions, sharing a post and sanctioning a received post. Agents in the simulation try to maximize their influence and popularity in the social network by sharing relevant content and providing appropriate sanctions. Accordingly, we define two goals for each agent—*Promoting Views* and *Social Acceptance*.

Promoting Views. All agents try to promote their views (POVs) by providing positive sanctions to what agrees with their views and negative sanctions to what does not.

Social Acceptance. All agents try to achieve social acceptance (i.e., higher aggregate positive sanctions) by changing their stance toward issues.

5.3.4 Simplifying Assumptions and Limitations

We make simplifying assumptions to operationalize user attributes and online sharing behavior. *First,* we assume views (on an issue) to be binary in this simulation, i.e., either supporting POV-1 or POV-2, meaning agents with no POV are nonparticipating. This is a design choice as

[1] https://github.com/ahaque2/MultiAgent_Social_Simulation

we intend to analyze the scenario where only motivated agents (i.e. agents who have a POV) try to influence and promote their views. As an agent becomes neutral in its POV (i.e., an agent with POV as zero), it stops sharing posts and providing sanctions. We assume all agents have some POV at the start of the simulation, and no agent has a neutral POV.

Second, we assume the initial user attributes and stance of each post based on a probability distribution. We use a random normal distribution to populate initial user attributes including the agent's stance towards an issue, sharing preference, and post's stance. This ensures a balance of stance toward each party across issues and provides a reasonable starting condition for the simulation.

Third, we assume all agents prefer getting positive sanctions over negative or none. They accordingly change their stance on issues over time to ensure social acceptance (i.e., to get aggregate positive sanctions from their neighbors). Sanctions also influence user activity; positive sanctions cause higher user activity while negative sanctions cause it to decline.

Our simulation models user preferences and emulates user behavior on social networks to analyze polarization dynamics. However, our model has a few limitations that stem from the simplifications (of user behavior and its influence).

First, for simplicity, sharing of posts and opinion shifts are sequential in this simulation, i.e., only one post is being shared in the network at any given time. Another post starts diffusing in the network only when the previous post has completely diffused (i.e., has reached all agents it could have). This limits the simulation to not factor in the effects of parallel exposure to different (maybe conflicting) information, i.e., being exposed to several posts relating to an issue before forming (shifting) an opinion about it.

Second, the social network in this simulation is static, i.e., neither a new link is formed nor an existing one severed at any time. However, selective exposure partially makes the network dynamic by filtering posts based on the difference in stance between two agents towards an issue. A dynamic social network demands far more computational resources and some knowledge of the offline world to link or delink agents over time appropriately.

5.4 Experiments and Results

We now describe the experimental setup and the metrics used to measure changes in the social network, followed by results.

5.4.1 Initial Simulation Setup

We use the Facebook social network from Leskovec and Mcauley (2012) to seed the simulation. The social network consists of 4039 nodes (agents) and 88 234 edges (neighbors) and an average clustering coefficient of 0.61.

The agents in the social network interact by sharing posts from a pool of artificially generated posts without replacement. The stance of the posts follows a bounded normal distribution (μ=0.00, $\sigma = 0.52$, min=−1, max=1) such that there is equal support and opposition for each issue. We predefine six issues and generate an equal number of posts for each issue. We use a total of ≈5000 posts shared between agents in each experiment run. Each simulation run ends when all posts in the pool of generated posts have been shared in the social network.

We create ten independent initial distributions to assign different initial user attributes for each simulation run. We set initial user satisfaction to zero for all agents. Each agent is initialized with a sharing preference based on a random normal distribution bounded between 0 and 1 (average over all distributions, μ=0.5, σ=0.14, min=0, and max=1). User activity is initialized based on a tailed distribution bounded between 0 and 1, skewed towards higher values (average over all distributions: μ=0.874, σ=0.17, min=0, and max=1). Higher initial user activity ensures greater activity and faster results. We compute kurtosis (Zwillinger and Kokoska 1999) for all user activity distributions. The average kurtosis (over all ten distributions of user activity) was 1.54 (for a normal distribution, kurtosis is zero).

We assume two POVS (Point-Of-Views), POV-1 and POV-2. Each agent has a POV in [−1, 1] that depends on its stance on various issues. Each agent's stance towards different issues is initialized based on a random normal distribution bounded in [−1, 1] centered around zero. The stance distribution is such that, on aggregate, there is equal support and opposition for each issue. The POV of each agent is computed as the average stance on issues favoring each POV resulting in a normal distribution in [−1, 1] approximately centered around zero (average over all distributions, μ =0.01, σ =0.11, min=-0.40, and max=0.44). This ensures approximately equal support for each POV at the start of the simulation.

We ensure consistency between the agent stance who authors and shares the post and the stance of the post by choosing the authors appropriately. If an agent supports issue A, it will only start a supportive post on issue A, whereas an agent who opposes it starts only a critical one. Agents are chosen to be authors of a post based on their activity score and sharing preference half of the time and at random for the other half. Agents who are more active or have a higher sharing preference are more likely to start sharing a post.

5.4.2 Evaluation Metrics

We define primary and secondary metrics to measure various changes in the network over time. Primary metrics measure polarization and user satisfaction, while secondary metrics compare initial and final user distribution for different user attributes for each experiment.

Primary Metrics

Primary metrics include the following.

Polarization. Polarization measures the extent to which the resulting distribution of opinions is polarized. We adopt the polarization index measure proposed by Morales et al. (2015) to measure overall polarization in the social network. The polarization index is inspired by the electric dipole moment and measures polarization as the distance between two opposing ideologies. Polarization lies in [0, 1] with 0 indicating least polarization and 1 indicating most.

To compute polarization We define A^{-} as the relative population with POV-1 (i.e., negative POV, <0) and A^{+} as the relative population with POV-2 (i.e., positive POV, >0). We compute the normalized population differences using the relative populations A^{-} and A^{+}.

$$\Delta A = |A^{+} - A^{-}| \tag{5.8}$$

We then compute each population's gravity center (mean), gc^{-}, and gc^{+}, and define the pole distance, d, as the normalized distance between the two gravity centers. d can be expressed as.

$$d = \frac{|\mathrm{gc}^{+} - \mathrm{gc}^{-}|}{|\max(A^{+}) - \min(A^{-})|} \tag{5.9}$$

$\max(A^{+})$ expresses the maximum possible value for positive opinions (POV>0), and $\min(A^{-})$ expresses the minimum possible value for negative opinions (POV<0).

The network polarization ($\mathrm{Polarization}(G, p_k)$) is defined based on the difference in size between the population of both POVs (ΔA) and the pole distance d.

$$\mathrm{Polarization}(G, p_k) = (1 - \Delta A) \times d \tag{5.10}$$

Polarity. Polarity is indicative of the POV that has greater aggregate support in the social network. We measure polarity as the mean POV of all agents. Polarity ranges over [−1, 1], with −1 indicating absolute support (by all agents) for one POV (POV-1) and 1 for the other (POV-2), and 0 for neutral POV.

$$\text{Polarity}(G, p_k) = \sum_{a_i \in G} \frac{\text{POV}(a_i, p_k)}{\text{numAgents}(G)} \tag{5.11}$$

Homophily. Homophily measures the homogeneity of a network structure with respect to some attribute (i.e., the agents' POV in this case). Homophily is shown to be useful in link prediction between users in a social network (Yuan et al. 2014). Higher homophily is indicative of greater segregation in a social network. We use the *assortativity* of a social network (Newman 2003) to measure homophily. The value of homophily ranges over $[-1, 1]$, with 1 indicating a perfectly assortative network and values in $[-1, 0]$ indicating a perfectly disassortative network.

$$\text{Homophily}(G, p_k) = \frac{\sum_i e_{\text{ij}} - \sum_i a_{\text{i}} b_{\text{j}}}{1 - \sum_i a_{\text{i}} b_{\text{j}}} \tag{5.12}$$

where e_{ij} is the fraction of edges in a network that connects a vertex of type i to one of type j, and a_i and b_j are the fractions of each type (based on the agents' POV) of the end of an edge attached to vertices of type i, and type j respectively. The type depends on the agent's POV, and we group agents into 20 equally spaced groups based on their POV. We use the networkx [2] implementation of assortativity to compute network homophily.

User Satisfaction. User satisfaction measures how satisfied the overall social network is based on the outcome of individual user actions. To operationalize the computation for user satisfaction (for each agent), we use the sanction scores that an agent gets for sharing posts with other agents in the social network to compute the update in user satisfaction (Equation 5.13). We take the mean of each user's satisfaction to compute overall network satisfaction (Equation 5.14).

$$\text{uSat}(a_x, p_k) = \text{uSat}(a_x, p_{k-1}) + c_2 \sum_{a_i \in \text{neighbor}(G, a_x, p_k)} \text{Sanc}(a_i, p_i, a_x) \tag{5.13}$$

$$\text{netSat}(G, p_k) = \sum_{a_i \in G} \frac{\text{uSat}(a_i, p_k)}{\text{numAgents}(G)} \tag{5.14}$$

where $\text{uSat}(a_x, p_k)$ refers to the user satisfaction of agent a_x after the post p_k has been shared, $\text{uSat}(a_x, p_{k-1})$ refers to the user satisfaction of agent a_x before the post p_k has

[2] https://networkx.org/documentation/stable/reference/algorithms/assortativity.html

been shared, and netSat(G, p_k) measures the overall network user satisfaction after post p_k has been shared.

Secondary Metrics

We define secondary metrics to compare user distribution (based on count) in the initial (at the start of the simulation run) and final (after completion of each simulation run) populations. We define three secondary metrics based on user attributes (such as user activity and user's POV) and the primary metric on user satisfaction. Secondary metrics are computed after all posts are shared. Secondary metrics include the following.

Satisfied users. User distribution (percentage) in initial and final populations with negative (<0), zero (=0), or positive (>0) user satisfaction.

Active users. User distribution (percentage) in initial and final populations with low (<0.75), medium (>0.75 and <0.90), or high (>0.90) user activity.

Polarized users User distribution (percentage) in initial and final populations with high (>0.10 or <—0.10) or low (>—0.10 and <0.10) intensity of POVs.

Table E.1 describes the secondary metrics and lists their thresholds.

5.4.3 Experiments and Results

To address $\text{RQ}_{\text{tolerance}}$ (*Does higher tolerance among users in a social network help mitigate polarization?*), we vary agents' tolerance levels. To address $\text{RQ}_{\text{exposure}}$ (*Does selective exposure to congenial information contribute to polarization?*), we vary the levels of selective exposure in our simulation. We analyze the influence of changing these configurations on the primary and secondary metrics.

To mitigate the effects of stochasticity, we ran the simulation ten times with different initial distributions for the agent's attributes while keeping the social network and shared posts the same to ensure a fair comparison. For each experiment, we compute the primary and secondary metrics. The reported results are averages of ten simulation runs.

Figures 5.1 and 5.2 compare how polarization, polarity, homophily, and user satisfaction change with more posts being shared under different experimental settings. Tables 5.1 and 5.2 summarize our findings for the two experiments. Tables 5.4 and 5.5 include results from the statistical analysis. Tables E.2 and E.1 describe the notation used to explain the simulation design and metrics, respectively.

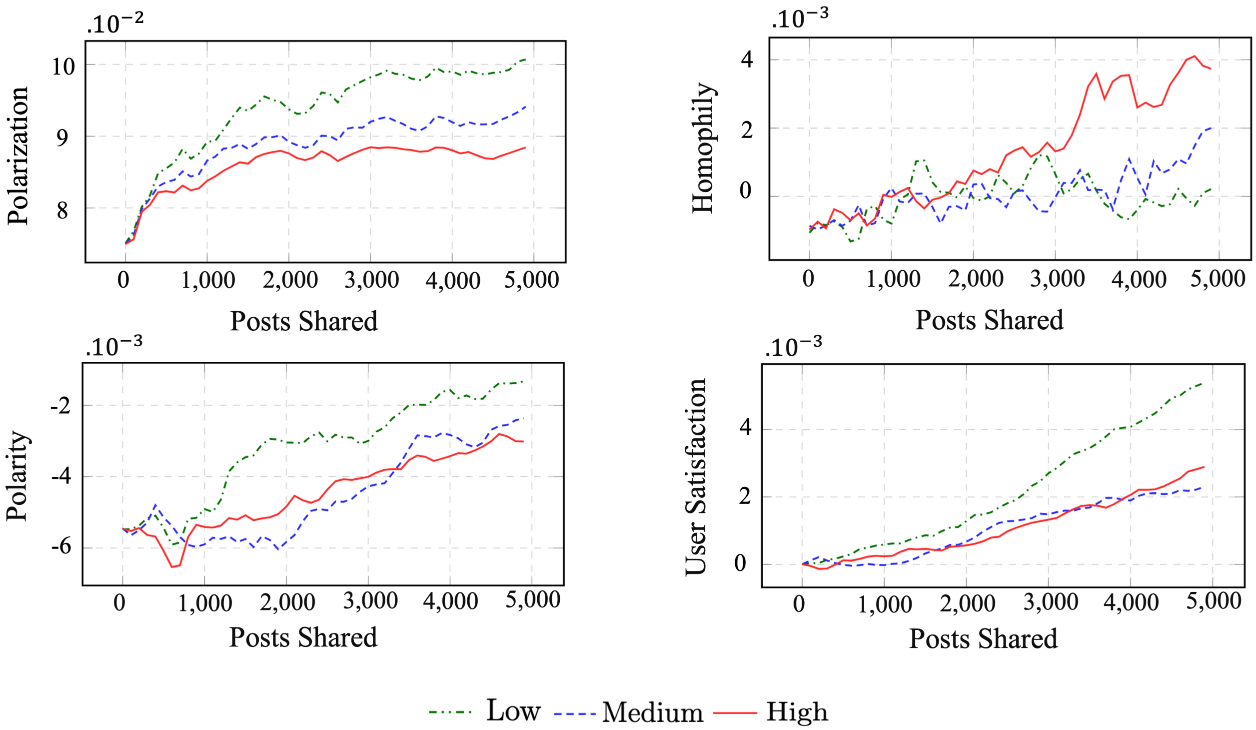


Figure 5.1: Experiment 1 (Tolerance): Varying user tolerance levels.

Experiment 1: Tolerant Users

The *tolerance* of an agent is defined based on its latitude of noncommitment (Sherif and Hovland 1961), i.e., the difference between the latitude of acceptance (assimilation) and latitude of rejection (contrast). The higher difference implies more tolerance. A more tolerant agent is less reactive to sanctions it receives from other agents for its shared posts, i.e., a more tolerant agent is less likely to change its stance on issues based on sanctions from agents who differ from its stance above a threshold (level of tolerance).

We run the simulation with three levels of tolerance, namely, HIGH, MEDIUM, and LOW. HIGH tolerant agents have a higher latitude of noncommitment (70%) and change their stance only based on sanctions from agents within a smaller (30%) difference in stance (between receiver and spreader) toward an issue. If a HIGH tolerant agent receives a sanction from an agent who differs in stance (on the issue in the shared post) greater than 30%, it discards that sanction and does not update its stance. MEDIUM tolerant agents have a latitude of noncommitment as 40%, and LOW tolerant agents have a latitude of noncommitment as 10%.

Figure 5.1 shows changes in the primary metrics as more and more posts are shared. When agents have a HIGH tolerance, polarization grows slower than when tolerance is MEDIUM or LOW. The polarization is constantly lower when tolerance in agents is HIGH compared to MEDIUM or LOW. Homophily grows faster when the agent's tolerance is HIGH, compared to MEDIUM or LOW, and social networks whose agents have higher tolerance end up with higher homophily after all posts are shared. The overall user satisfaction at LOW tolerance is constantly higher than HIGH or MEDIUM.

Table 5.1 shows the proportion of receiver (spreader and disinterested) and nonreceiver agents after all posts are shared. The number of receivers (agents who receive a post) is highest when tolerance is MEDIUM and lowest when tolerance is HIGH. The number of disinterested agents is highest when tolerance is HIGH.

Table 5.2 lists values for secondary metrics after all posts are shared. Secondary metrics compare the proportion of satisfied, active, and polarized users in the initial (before any posts are shared) and final (after sharing 5000 posts) populations based on thresholds defined for secondary metrics (Table E.1). The number of positively satisfied users is highest when tolerance in users is HIGH and lowest when tolerance is LOW. User activity shows minor variation across different levels of tolerance. LOW tolerance leads to the highest increase in highly polarized agents, whereas it is lowest when tolerance in agents is HIGH.

Finding 1: Tolerance

Higher tolerance in users slows down polarization, leading to a less polarized network, higher network homophily, lower user satisfaction, and a lower number of highly polarized users than when tolerance in users is lower.

Experiment 2: Selective Exposure

We emulate selective exposure in our simulation by exposing each agent only to posts from other agents who have a similar stance on the issue discussed in the post. To operationalize selective exposure, we use a threshold value of the difference in the stance of two agents beyond which they stop seeing each other's posts. An agent sees posts only from other agents whose stance differs from its stance on an issue in the post below a threshold. We experiment with four threshold values for selective exposure, NONE (allow all agents to see all content shared by neighbors without any filtering, i.e., no selective exposure), LOW (allow a difference of 80% in the stance between sharing and receiving agents toward the issue in the post), MEDIUM (allow 50% difference), and HIGH (allow 20% difference). We maintain the level of tolerance among users at MEDIUM for all scenarios in this experiment.

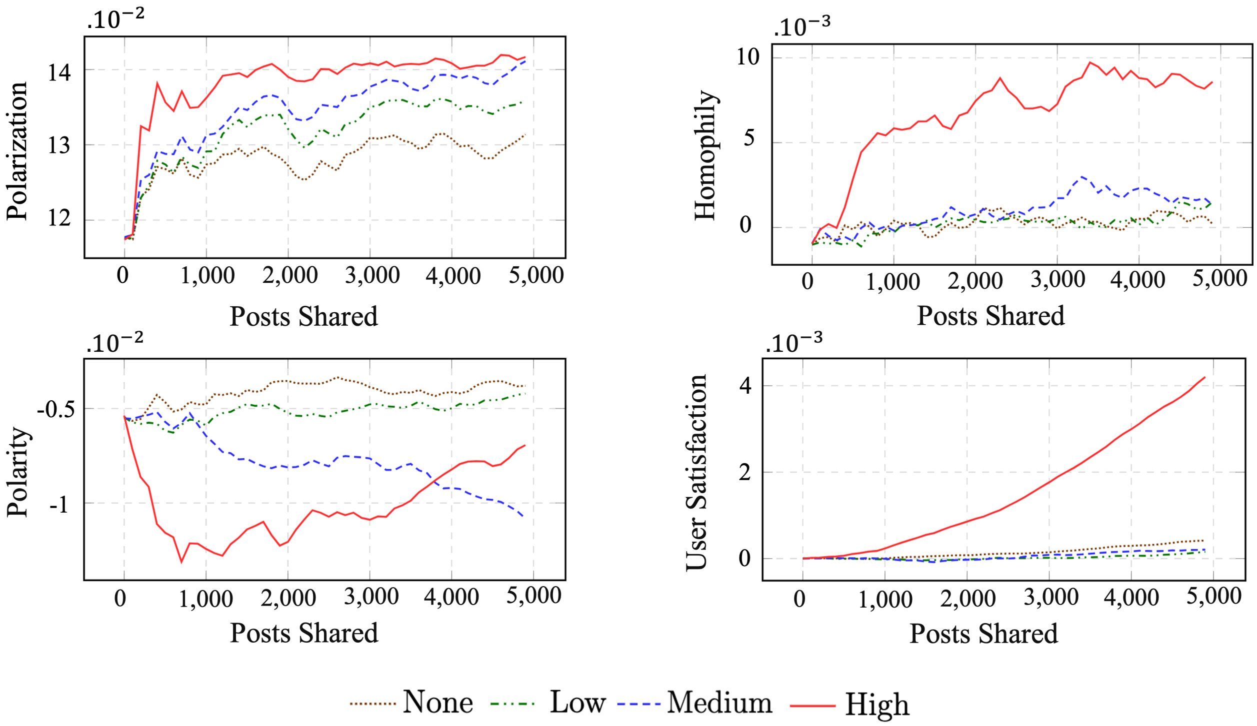


Figure 5.2: Experiment 2 (Selective Exposure): Varying selective exposure.

Figure 5.2 compares the influence of different levels of selective exposure on all primary metrics. HIGH selective exposure leads to the highest polarization, and NONE leads to the lowest. Polarization in a social network is constantly higher for higher levels of selective exposure. Homophily is highest when selective exposure is HIGH, and shows minor variations across lowers levels of selective exposure. User satisfaction is highest when selective exposure is HIGH and shows minor differences across lower levels of selective exposure.

Table 5.1 shows the proportion of receiver (spreader and disinterested) and nonreceiver agents after all posts are shared. HIGH selective exposure experiences the lowest proportion of receiver agents, while NONE selective exposure leads to most.

Table 5.2 compares the proportion of satisfied, active, and polarized users in the initial (before any posts are shared) and final (after sharing 5000 posts) populations based on thresholds defined for secondary metrics Table E.1. MEDIUM selective exposure experiences the highest number of positively satisfied users, whereas the highest number of negatively satisfied users is with NONE selective exposure. HIGH selective exposure leads to the lowest number of negatively satisfied users. The number of highly active users experiences the most decline when selective exposure is MEDIUM, and the least when selective exposure is HIGH. HIGH selective exposure leads to the highest number of highly polarized users, whereas NONE and LOW selective exposure leads to the lowest.

Finding 2: Selective Exposure

Higher selective exposure leads to higher polarization, higher network homophily, higher overall user satisfaction, and a higher number of polarized users than when selective exposure is lower.

Exp	**Config**	**Agent State**		
			Receiver	
		Nonreceiver	Spreader	Disinterested
Tolerant Users	LOW	60.12	14.49	25.39
	MEDIUM	53.95	17.30	28.75
	HIGH	62.99	13.36	23.65
Selective Exposure	NONE	54.76	16.88	28.36
	LOW	55.44	16.48	28.08
	MEDIUM	58.90	13.80	27.30
	HIGH	82.63	4.97	12.40

Table 5.1: Distribution of agents in the final population (Values in % of the total population).

Statistical Analysis

We conduct statistical analysis to test if different levels of selective exposure and tolerance lead to statistically significant differences in users' POV (point-of-view) and primary metrics (network polarization, homophily, polarity, and user satisfaction). For users' POV, we compare the final distributions (after all posts are shared) of users' POV at different levels of selective exposure and tolerance to establish if the differences are significant. For primary metrics, we compare the distributions of each primary metric (computed after sharing each post) at different levels of selective exposure and tolerance to identify differences in the overall social network metrics.

To choose the applicable statistical tests appropriately, we first evaluate the distributions. We test the normality of distribution using the Shapiro-Wilks normality test (Shapiro and Wilk

Exp	Config	User Satisfaction			User Activity			User Polarity	
		Neg	Zero	Pos	Low	Medium	High	Low	High
Initial Distribution		0.00	100.00	0.00	1.56	64.82	33.62	99.33	0.67
Tolerant Users	Low	52.09	23.79	24.11	4.51	64.79	30.70	97.50	2.50
	Medium	51.92	22.26	25.82	4.51	64.52	30.97	98.19	1.81
	High	50.11	23.42	26.47	4.33	65.39	30.28	98.54	1.46
Selective Exposure	None	51.74	23.25	24.01	4.48	64.82	30.70	98.69	1.31
	Low	51.40	22.93	25.67	5.08	64.62	30.30	98.69	1.31
	Medium	45.26	26.64	28.10	7.30	64.42	28.28	97.03	2.97
	High	23.75	53.97	22.28	4.35	63.98	31.67	96.73	3.27

Table 5.2: Comparison between initial and final distributions of agents (values in % of the total population).

1965). We use parametric statistical tests, namely paired t-test and one-way ANOVA, to compare normal distributions and nonparametric tests, namely the Kruskal-Wallis test, for distributions that are not normal.

In addition to the statistical significance test, we compute each test's effect size. For parametric statistical tests, we use Cohen's d (Cohen 1988) to compute the effect size as the distributions under comparison have similar standard deviations, and the sample size is large (≈4000). To interpret the effect size computed using Cohen's d, we adapt the interpretation from Cohen (1988) (see Table 5.3). For nonparametric statistical tests (Kruskal-Wallis test), we use epsilon square (ϵ^2) (Kelley 1935) to compute the effect size based on recommendations from Tomczak and Tomczak (2014). To interpret the effect size computed using epsilon square (ϵ^2) we adapt the interpretation from Rea and Parker (2014) for the correlation coefficient and square threshold values of each bin because ϵ^2 is a squared metric. The resulting interpretation for ϵ^2 effect size we use is as shown in Table 5.3. we chose ϵ^2 over other popular alternatives such as omega-squared (ω^2) (Albers and Lakens 2018), because ϵ^2 is less biased (Okada 2013).

For all statistical significance tests, we assume the null hypothesis indicates a similar distribution between compared entities. In contrast, the alternate hypothesis indicates that statistically significant differences exist in the compared distributions.

We use the significance level, i.e., alpha as 0.01, to accept or reject the null hypothesis.

We use the Kruskal-Wallis test to compare all primary metrics for different levels of selective exposure and user tolerance. For selective exposure, we compare how different levels (i.e., LOW, MEDIUM, and HIGH) compare against NONE selective exposure. In contrast, for user tolerance, we compare each level of tolerance against each other in pairs.

Table 5.4 shows the results of the statistical significance test for all primary metrics at different levels of selective exposure and tolerance. The compared distributions correspond to the value of each metric after each post is shared on the social network, effectively comparing how the social network evolves (in terms of the metrics) as more and more posts are shared. The p-values for each pair of distributions comparing the metrics indicate that the difference in the distributions is statistically significant, and the null hypothesis can be rejected, though the effect sizes vary. Based on the effect size, the difference between network homophily when selective exposure is MEDIUM and HIGH (compared to NONE selective exposure) is very strong. The difference in polarization at HIGH selective exposure (compared with NONE) and high tolerance (compared with LOW) is strong. Similarly, the difference in homophily between LOW and NONE selective exposure and user satisfaction between HIGH and NONE selective exposure is also strong. For different levels of user tolerance, relatively strong differences exist in polarization between LOW and MEDIUM, MEDIUM and HIGH; in homophily between LOW and MEDIUM, HIGH and LOW; and in polarity between HIGH and LOW. For different levels of selective exposure, a relatively strong difference (in comparison to NONE selective exposure) exists in polarization at HIGH; in polarity at MEDIUM; and in user satisfaction at LOW. Other comparisons have an effect size of either moderate or weak.

Table 5.5 shows the results of the statistical significance test comparing users' POV at different levels of selective exposure and tolerance. The compared distributions correspond to each user's POV after all posts are shared on the social network, effectively comparing how the POV of users differ as a consequence of different levels of selective exposure and tolerance at the start and end of each simulation run. The p-values for some of the differences show that the differences are statistically significant, though the effect sizes are either small or very small.

	Effect Size	**Interpretation**
Epsilon-Square (ϵ^2) Interpretation based on (Rea and Parker 2014)	[0.00, 0.01)	Negligible
	[0.01, 0.04)	Weak
	[0.04, 0.16)	Moderate
	[0.16, 0.36)	Relatively strong
	[0.36, 0.64)	Strong
	[0.64, 1.00]	Very strong
Cohen's d (Cohen 1988)	0.20	Small
	0.50	Medium
	0.80	Large

Table 5.3: Effect sizes and their interpretations (according to the cited works).

Exp	Metric	Dist1	Dist2	H-statistic	p-value	Effect Size
Tolerant Users	Polarization	Low	Medium	2784.62	<0.01	0.27
		Medium	High	2852.45	<0.01	0.28
		High	Low	4178.42	<0.01	0.42
	Homophily	Low	Medium	1894.71	<0.01	0.19
		Medium	High	15.27	<0.01	0.00
		High	Low	2353.32	<0.01	0.24
	Polarity	Low	Medium	67.88	<0.01	0.01
		Medium	High	1516.77	<0.01	0.15
		High	Low	1981.18	<0.01	0.20
	User Satisfaction	Low	Medium	1111.50	<0.01	0.11
		Medium	High	10.60	<0.01	0.00
		High	Low	1075.30	<0.01	0.11
Selective Exposure	Polarization	None	Low	1336.62	<0.01	0.13
		None	Medium	2918.22	<0.01	0.29
		None	High	4317.15	<0.01	0.43
	Homophily	None	Low	5038.38	=0.04	0.50
		None	Medium	7316.85	<0.01	0.73
		None	High	7485.42	<0.01	0.75
	Polarity	None	Low	4.00	<0.01	0.00
		None	Medium	1813.12	<0.01	0.18
		None	High	6349.25	<0.01	0.63
	User Satisfaction	None	Low	2927.38	<0.01	0.29
		None	Medium	1232.89	<0.01	0.12
		None	High	4286.36	<0.01	0.42

Table 5.4: Statistical significance test results.

5.5 Findings

Polarization is slowed down substantially when tolerance in users is HIGH. HIGH tolerant users experience the least network polarization and have less network polarity than when users' tolerance is LOW. The low polarization is plausibly because HIGH tolerant users are less likely to change their stance on issues based on sanctions they receive than LOW tolerant users, hence, slowing down change to a users' POV. The number of highly polarized users is lowest when user tolerance is HIGH. Our results are consistent with the earlier work (Coscia and Rossi 2022),

Exp	Test	Dist1	Dist2	Test Statistic	p-value	Effect Size
Tolerant Users	Paired t-test	Low	Medium	1.35	=0.18	0.02
		Medium	High	0.72	=0.47	0.01
		High	Low	2.06	=0.04	0.03
	One-way ANOVA	Low	Medium	1.08	=0.30	0.02
		Medium	High	0.26	=0.61	0.01
		High	Low	2.41	=0.12	0.03
Selective Exposure	Paired t-test	None	Low	1.03	=0.30	0.02
		None	Medium	10.20	<0.01	0.17
		None	High	3.99	<0.01	0.07
	One-way ANOVA	None	Low	0.56	=0.45	0.02
		None	Medium	57.66	<0.01	0.17
		None	High	9.48	<0.01	0.07

Table 5.5: Statistical significance test comparing initial and final POV distributions.

which found lower levels of network polarization with high user tolerance in a social network.

Figure 5.1 shows that user satisfaction is constantly higher when tolerance is LOW. Consequently, low user tolerance leads to higher overall user satisfaction. However, the number of users with positive satisfaction is higher when tolerance is HIGH, compared to when tolerance is MEDIUM or LOW (Table 5.2). This indicates that the sharing of posts in a social network whose users have lower tolerance leads to higher overall user satisfaction but concentrated among fewer users.

Surprisingly, HIGH user tolerance leads to a more homophilic network (based on users' POV) than when user tolerance is LOW or MEDIUM. Also, user reach (number of users who receive a post) is lower when tolerance in users is HIGH compared to LOW and MEDIUM.

HIGH selective exposure leads to higher polarization than MEDIUM, LOW, and NONE selective exposure, in that order. This is plausibly because when selective exposure is HIGH users are more likely to see congenial posts (posts that agree with their existing stance) and are subject to fewer posts that may challenge their stance. Our finding that higher selective exposure leads to higher polarization agrees with earlier findings from prior works (Stroud 2010; Garrett et al. 2014; Kim and Chen 2016). However, it is important to elucidate the difference in the methodology between our work and prior works to understand the results better. While ours is a multiagent simulation that captures the evolution of polarization as caused by the social interactions between users, prior works (Stroud 2010; Garrett et al. 2014; Kim and Chen 2016) primarily rely on self-reported survey data for their conclusions. Further, prior works focus on how exposure to some information may polarize an individual's attitude in isolation rather than as a consequence of social interactions between multiple users.

As expected, user satisfaction is higher for higher levels of selective exposure (Figure 5.2). High user satisfaction may result because users receive more congenial posts with higher selective exposure, leading to more positive sanctions and higher user satisfaction for some users. The number of users with zero user satisfaction (i.e., users whose satisfaction didn't change during the simulation run) is highest when selective exposure is HIGH and the number of negatively satisfied users is substantially lower ($\approx 2\times$) than lower levels of selective exposure. This indicates selective exposure ensures fewer users end up with aggregate negative satisfaction.

Higher selective exposure leads to the lowest user reach (i.e., the highest number of non-receivers, Table 5.1). This is most likely due to filtering out noncongenial posts for each user, which leads to fewer users receiving any given post than when no selective exposure is applied. The number of disinterested is lowest in the case of HIGH selective exposure, demonstrating that selective exposure makes it less likely for a post to reach potentially disinterested (i.e., users with a potentially noncongenial POV toward the post). This comes at the cost of a low number of spreaders when selective exposure is HIGH.

HIGH selective exposure witnesses the least drop in highly active users between the start and the end of the simulation. Our findings on higher selective exposure leading to more highly active users are consistent with some empirical findings from prior work. Prior work (Stroud 2010) found selective exposure to congenial political information increases participation. At the same time, it undermines earlier work that found a positive role of cross-cutting exposure on political participation (Kim and Chen 2016).

HIGH selective exposure leads to the highest number of highly polarized users at the end of the simulation. HIGH selective exposure also leads to a social network with the highest homophily. Homophily shows some of the highest effect sizes in the statistical significance test analysis with values indicating a very strong relation implying that the change in overall network homophily is statistically significant. The effect size is highest when selective exposure is HIGH, followed by MEDIUM, and then LOW indicating an increasing pattern of homophily with higher selective exposure.

Our findings have practical and valuable implications for social networking platforms that have become an integral part of our lives. These platforms try to maximize user satisfaction and often employ content filtering (algorithm selective exposure) to choose content based on user preference. Our simulation shows achieving user satisfaction via selective exposure can potentially increase polarization in the social network. High selectivity in exposure to congenial content may lead to better user satisfaction (due to increased likelihood of viewing congenial posts), but it also leads to more polarized users. On the other hand, social networks whose users have a higher tolerance experience far less polarization among their users for the same number of shared posts. However, the user satisfaction when users' tolerance is higher is lower.

Interestingly, network homophily (the tendency of being connected to users with similar POV) increases in both experiments, i.e., higher selective exposure and higher tolerance in users both lead to networks with higher homophily. Social networks with higher homophily are more prone to forming echo chambers (wherein a person encounters only beliefs or opinions that coincide with their own), which is a growing challenge for social media platforms. While it is not incumbent on social networking platforms to mitigate its ill effects, such as polarization among users and the formation of echo chambers, there are some benefits to it. For instance, our simulation shows higher selective exposure leads to the lowest user reach (i.e., highest number of nonreceivers).

Our simulation model is a step toward understanding the social interactions between users in a social network and how it influences user behavior and polarization. A better understanding the potential consequences of the interactions on a social network can show us ways to mitigate the ill effects while still making the most of these social networking platforms.

5.6 Threats to Validity

Modeling user behavior is a challenging task that demands an intricate understanding of human psychology and an extensive operationalization of human traits. Though we model each user based on theories from social science and relevant observations from previous related works, the simplifications done to formalize the setup incur some threats to validity.

First, we assume equal strength of ties between each pair of connected users. In reality, people have varying strengths of ties, affecting how they react to posts from others and how it influences them.

Second, we only consider a user's own preferences and content of the post when deciding to share a post, and providing sanctions. In reality, there may be a myriad of factors that affect such decisions.

Third, the simulation runs on artificially generated data. User attributes and the posts being shared are artificially generated based on suitable probability distributions. Though we ensured appropriate distributions for initial user attributes, this does not guarantee a reasonable replication of a real-world social network. Any generalizations based on these findings need to be verified with empirical data.

Forth, the results are based on simulation runs each of which ends after sharing ≈ 5000 posts. While most plots indicate the simulation stabilizing (near the end of the simulation) with the general direction of the plots being stable, there is no certainty that the same trends will continue forever.

The results should be taken with caution. Although our model is based on assumptions grounded in prior studies on polarization on social media, we use artificially generated data for this analysis. Further, reliably modeling user behavior is nontrivial and requires a fine-grained understanding of user behavior. We make simplifying assumptions in our model.

5.7 Future Directions

This work brings forth exciting directions for further research. First, it would help to develop richer simulation models that capture the dynamics of social networks, such as forming and severing ties between users and diffusing several posts simultaneously in the network. We could not incorporate such dynamism in this work due to the computational overhead. Second, instead of seeding the simulation with artificially generated data, it would be useful to run the same study on real data collected via a human-subject study. However, getting such data is challenging due to privacy concerns that such data collection raises. Third, we would like to extend this work to directional social networks. The current simulation uses a bidirectional Facebook social network, i.e., if agent A is friends with agent B, it implies that agent B is also friends with agent A. However, many real-world social networks are unidirectional, i.e., a link from agent A to agent B doesn't imply a link in the reverse direction. For instance, you can follow someone on Twitter without them following you back. It would be interesting to extend this simulation to such use cases and compare it with the results from the current setup. Fourth, it would be useful to extend our model to incorporate methods of intervention that can help mitigate polarization in a social network.

CHAPTER

6

CONCLUSION

This dissertation attempts to create a computational framework to analyze political news and social media discourse related to it, and evaluates the effects of selective exposure to congenial (like-minded) content on polarization dynamics in a social network using a multiagent social simulation. We integrate social science theories – including moral foundation theory (Graham et al. 2009; Roy and Goldwasser 2021), war and peace journalism framework (Galtung 1986, 1998), and social judgment theory (Sherif and Hovland 1961) – with computational modeling using pretrained language models and multiagent social simulation.

Additionally, we contribute two open-source datasets to support news and social media analysis. One dataset includes online political news and user responses on social media, enabling the study of political slants in news and variations in user engagement on social media. The other dataset includes conflict-related news manually annotated using the war and peace journalism framework and can be used to train computational models to analyze conflict-related news.

6.1 Summary of Findings

Our research reveals several key findings, such as political slant in the news, variations in reader reactions on social media, agenda-setting in war news, and the effects of selective exposure on polarization dynamics in a social network. The following sections summarize these insights.

6.1.1 Political Slants in News and Moral Divide in User Response

Election-related news coverage exhibits political slants in both traditional online news and social media content, with significant differences across publishers with varying (presumed) political leanings. Moreover, social media features more sentiment-driven content than traditional online news, and political slant is more pronounced in news shared on these platforms.

A moral analysis of user responses to news on social media reveals a concerning trend: readers' moral reasoning varies significantly across different news sources, suggesting partisan influence. Moreover, these differences show a variation across social and political issues.

These findings raise concerns about the media's role in reinforcing ideological divides, the strategic use of social media to advance political agendas, and the existence of a moral divide among audiences of different news sources on social media.

6.1.2 Moral Analysis of War News

The moral analysis of war news reveals significant variations in how different news publishers frame the same events. Publishers report on conflicting entities using distinct moral perspectives, with these differences aligning with their presumed (based on prior research) political leanings. This pattern suggests a partisan coverage of war news.

A notable trend is the causal relationship between news coverage across different publishers. This suggests that the moral framing used by one news publisher influences the framing adopted by competing publishers. Such influence aligns with agenda-setting, where one publisher's reporting shapes the overall narrative.

We found some evidence of cross-entity causal associations across news publishers, which implies that the way one publisher frames an entity in the news influences how a competing publisher frames a conflicting entity. This suggests a potential reactionary pattern in news coverage, where one publisher's reporting prompts a counter-narrative from another. Whereas causal associations in the moral framing of the same entity may be driven by external events or confounding factors, cross-entity causal associations are more complex and concerning.

6.1.3 War and Peace Journalism

An analysis of war news coverage through the lens of the war and peace journalism framework reveals signs of partisan reporting among news publishers. Each publisher shapes narratives to reflect its ideological stance, casting different entities as either villains or victims of the same news events.

Furthermore, the application of war and peace frames differs across publishers. However, despite their political leanings, all news publishers incorporate elements of both frames in their coverage.

Our analysis identifies limitations of the war and peace journalism framework in practice. The presence of a peace frame does not necessarily indicate peace-supportive coverage, nor does a war frame inherently promote war-supportive narratives. Additional context is required to determine whether a given frame truly aligns with peace-supportive or war-supportive news coverage. This is because the use of war or peace frames is often driven by a publisher's perception of newsworthiness for its audience rather than a deliberate effort to advocate for peace or war. However, beyond the selection of frames, the way news is presented is crucial in shaping its interpretation.

6.1.4 Dynamics of Polarization in Social Networks

An analysis of polarization dynamics in a social network reveals that while selective exposure to congenial (like-minded) content enhances user satisfaction by tailoring content to individual preferences, it contributes to ideological divides, intensifying polarization within the network. This dual effect is key to understanding how content recommendation systems, which selectively expose users to congenial content to enhance user satisfaction, shape public discourse, and influence engagement on online social platforms.

The findings suggest that increased user tolerance (toward noncongenial content) can help mitigate polarization; however, this often comes at the expense of user satisfaction, as exposure to opposing viewpoints may be perceived as uncomfortable or displeasing. Conversely, selective exposure to congenial (like-minded) content enhances user satisfaction but simultaneously intensifies polarization and limits the reach of content across diverse audiences.

6.2 Research Significance and Social Impact

6.2.1 Partisan Political News and Social Media Divide

News coverage in recent times has become increasingly polarized (Chinn et al. 2020). As a consequence of partisan news, polarization is on the rise (Boxell et al. 2020; Dimock and Wike 2021), and the gap between the political values between opposing views is now wider than at any point in recent history (Heltzel and Laurin 2020; Pew 2017). Moreover, many believe the news is biased (Jones 2018), and the trust in news media is at a decline (Survey 2018). This widening divide among people and declining trust in media hinders constructive discussions on important social and political issues, highlighting the need for further research on news media and social media discourse surrounding news.

This dissertation enhances our understanding of political slants in news and their impact on social media users by developing computational frameworks and resources for such analysis. By developing the NEWSSLANT dataset, which encompasses over 36,000 news headlines and approximately 4 million reader reactions across various media outlets, this study provides a robust resource for analyzing the dynamics of media influence on political discourse, particularly in the context of the 2020 US presidential elections. To the best of our knowledge, this is the first dataset that includes election-related news from traditional online platforms and its parallel coverage on social media, along with reader reactions to the news.

A key finding of this research is the disparity in political slants between online news published on traditional platforms and the same publishers' content on social media. This indicates that news publishers strategically leverage social media to promote their agendas, as evidenced by a stronger political bias and greater alignment with their ideological stance in social media news compared to the same publishers' content on traditional online news platforms. This is especially concerning, given the use of social media as a tool to get news is on the rise, particularly among young adults (Infield 2020).

User responses on social media show a moral divide among audiences of different news sources. Response to the same news events differs in its moral focus. For example, reactions to coverage of Black Lives Matter (BLM) protests varied sharply; audiences of right-leaning publishers emphasized values such as sanctity and degradation, whereas those of left-leaning publishers prioritized harm, fairness, and betrayal. Similar patterns emerged in responses to other key social and political issues, including immigration, healthcare, and the economy. This divergence in moral foundations suggests that followers of different news sources focus on distinct aspects of the same issues, potentially reinforcing ideological divides. This is especially concerning as studies suggest a correlation between online and offline political participation (Bode et al. 2014).

This study contributes to the ongoing discussions on the ethical responsibilities of news publishers and the significance of balanced reporting in fostering constructive public dialogue. The findings highlight the urgent need for media literacy initiatives that raise awareness of how news framing shapes perceptions and opinions, particularly on social media platforms. This work provides a valuable foundation for future research focused on fostering non-partisan online public discourse and improving the quality of news reporting in an increasingly fragmented media landscape.

6.2.2 War and Peace Journalism and Moral Framing in War News

News is often the only source of information about the developments in an ongoing war. As such, it shapes public perception of the conflict and can, in turn, influence its trajectory (Hallin 1989; Kull et al. 2003; Aday et al. 2005). However, war news makes it challenging to maintain journalistic objectivity, particularly if the journalist's own country is involved (Bläsi 2009). Moreover, war news reports about inherently charged situations rife with events of high emotional power. This incentivizes news publishers to produce emotionally charged content, as it drives higher user engagement and increased sharing online (Pröllochs et al. 2021b,a; Brady et al. 2017; Berger and Milkman 2012; Chuai and Zhao 2022). This underscores the need for a critical analysis of war coverage and evaluation of the framing across news publishers.

This dissertation presents a computational approach to analyzing war news through a lens of war and peace journalism and moral framing. The findings reveal significant differences across news publishers' moral framing and use of war and peace frames. Moreover, the portrayal of victims and villains in war news coverage across publishers indicates potential partisan biases in news reporting.

Our findings uncover an alarming trend: the causal relationship between news publishers' moral framing. The findings indicate that a publisher's moral framing influences the moral framing adopted by competing publishers, suggesting an agenda-setting effect. More concerning is that this influence extends across conflicted entities, suggesting a potentially reactionary news coverage, where news coverage by one publisher about a conflicted entity influences how a competing publisher frames news about another conflicted entity.

Additionally, our analysis uncovers the limitations of war and peace journalism in practice. Specifically, the presence of a peace frame does not necessarily equate to peace-supportive coverage, nor does the presence of a war frame guarantee war-supportive coverage. Determining whether a headline aligns with war or peace-supportive coverage requires additional context. Factors such as a publisher's target audience and geographical location are crucial in shaping news coverage. For example, a news publisher reporting from a region distant from the war zone may emphasize the invisible (indirect) and broader consequences of a conflict (both peace frames), not necessarily to adopt a peace frame, but because such coverage aligns with the perceived newsworthiness of its audience. In contrast, a publisher located within or near the conflict zone may focus on war-related frames, and report on immediate and visible effects of war to provide updates focusing on the war context, not necessarily to promote a war-supportive narrative but because these aspects directly affect its audience.

This research contributes to the growing body of literature on war news analysis. Unlike many prior studies that assess moral framing through a binary lens of right and wrong, a subjective approach that often lacks consensus, this study presents a multidimensional analysis of moral framing using moral foundations theory. This can help in advancing research on computational approaches aimed at delivering news with a more balanced moral framing, particularly in the context of war reporting. Our findings highlight the need for a more nuanced analysis of news and identify challenges and pitfalls for future research to address.

6.2.3 Selective Exposure and Polarization in Social Networks

The exponential growth of information presents a computational problem of identifying effective ways of finding relevant content for end users. Recommendation systems are designed for exactly this task. They enable navigating online content and facilitate users to find relevant content without getting lost in the sea of available content. They also provide a pleasant experience by personalizing content for each user.

A challenge for recommendation systems is to identify personalized recommendations for each user. Many recommendation systems are designed to optimize for metrics like user satisfaction and engagement, which, while reasonable, can lead to unintended adverse consequences. This is because individuals are more likely to engage with information that aligns with their preexisting beliefs (Klapper 1960; Redlawsk 2002; Taber and Lodge 2006). This could cause people to gravitate toward information that aligns with (and confirms) their existing views, leading to the formation of echo chambers and filter bubbles (Pariser 2011; Cinelli et al. 2021; Nguyen 2020).

The ability to connect with like-minded users and filter content on social media, coupled with selective exposure to personalized content, could potentially deepen the existing divide in news exposure across different audiences. For example, Facebook's content recommendation algorithms, optimized for engagement rather than content quality, have been linked to the spread of ethnic conflicts, such as the Rohingya conflict (Ansar and Khaled 2023; Stein Tønnesson and Aung 2022; Tähtinen 2024). Much like newsroom editors, these algorithms influence which stories gain visibility, thereby playing a crucial role in shaping public discourse. Yet, despite their substantial influence, they operate with minimal accountability, prompting serious concerns about transparency and oversight.

This research provides a computational approach to analyze how selective exposure to congenial (like-minded) content affects polarization dynamics in a social network. We analyze how individual behaviors and interactions contribute to generating and escalating polarization based on information sharing in a social network. One of the critical contributions of this research is its exploration of the trade-offs associated with content personalization. The findings reveal that while the content filtering algorithms designed to cater to user preferences can enhance individual satisfaction, they inadvertently foster ideological divides, leading to greater polarization in a social network. This duality is crucial for understanding how content recommendation practices can shape public discourse and influence user engagement in online social networks.

Furthermore, the research emphasizes the role of tolerance and selective exposure on polarization dynamics and users' satisfaction in a social network. The findings indicate that higher user tolerance can mitigate polarization, but this often comes at the cost of lower user satisfaction. Conversely, selective exposure to congenial content amplifies polarization while reducing the reach of content due to increased content filtering.

This research contributes to the broader discourse on the implications of social media in public discourse and political engagement, offering a computational framework that can be utilized for further research in understanding and addressing polarization in online communities. The insights from this research are vital for informing social media platforms and content moderation strategies, emphasizing the need for a balanced approach to content recommendation that fosters diverse perspectives and mitigates online polarization.

6.3 Future Directions

6.3.1 Identifying 'Intermedia' Agenda-Setting

Although the current computational analysis presents useful findings, it is limited to a few news publishers, primarily US-based, Aljazeera (Qatar), and the BBC (UK). Moreover, the analysis is limited to text news only. In future work, we would like to apply our computational framework to analyze diverse types of news media, such as print (newspapers), visual (television news), and audio (radio and podcasts).

We are particularly interested in investigating 'intermedia' agenda-setting, i.e., the influence of one media over the other. For instance, does print media influence television news, or vice versa, or does it show no correlation?

As multimedia platforms play an increasingly significant role in news dissemination and public perception, expanding the analysis to diverse media formats may uncover meaningful differences in how news is framed and consumed across various media.

6.3.2 Expanding Beyond Moral Framing In News

The current approach identifies moral vector subspace by employing a semi-supervised vector subspace projection approach (Bolukbasi et al. 2016; Field and Tsvetkov 2019). A key advantage of this approach is that it does not require a labeled training dataset; instead, it relies on word pairs to define a word embedding vector subspace for a given concept. Currently, we use word pairs derived from the enhanced Moral Foundations Dictionary (eMFD) to isolate the moral vector subspace, but this approach could be extended to identify other vector subspaces in the contextual representation of text.

Earlier methods have used this approach to identify gender subspace (Bolukbasi et al. 2016), sentiment subspace (Field and Tsvetkov 2019), and moral directions from right and wrong actions (Schramowski et al. 2022). As a future work, we aim to extend this approach to identify vector subspaces for dimensions related to the ideological framing of news.

6.3.3 Using LLM Agents to Mimic Human Behavior

The agents in the current social simulation have limited capabilities and lack sophistication. In future work, we aim to enhance them by incorporating Large Language Models (LLMs) as agents. Incorporating LLM agents would enable the evaluation of a much wider range of scenarios that the current method struggles to operationalize. For example, LLMs can mimic human behavior through simple prompts, allowing for more realistic agent behavior. However, replacing existing simple agents with LLM-driven agents would substantially increase the computational cost of running the simulations.

Additionally, we would like to experiment with other social network structures. The current simulation is based on a bidirectional Facebook-style social network, where a friendship between A and B exists only if both users are mutually connected. However, different social network structures may yield different observations due to their inherent properties. For example, Twitter and Instagram follow a unidirectional structure, allowing A to follow B without requiring B to follow A in return. In future work, we aim to extend the simulation to incorporate a variety of social network structures to better understand how these differences influence the dynamics of information spread and user interactions.

REFERENCES

Aday, S., Livingston, S., and Hebert, M. (2005). Embedding the truth: A cross-cultural analysis of objectivity and television coverage of the iraq war. *Harvard International Journal of Press/Politics*, 10(1):3–21.

Ahmad, T., Alvi, A., and Ittefaq, M. (2019). The use of social media on political participation among university students: An analysis of survey results from rural Pakistan. *SAGE Open*, 9(3):1–9.

Albanese, F., Pinto, S., Semeshenko, V., and Balenzuela, P. (2020). Analyzing mass media influence using natural language processing and time series analysis. *Journal of Physics: Complexity*, 1(2):025005.

Albers, C. and Lakens, D. (2018). When power analyses based on pilot data are biased: Inaccurate effect size estimators and follow-up bias. *Journal of Experimental Social Psychology*, 74:187–195.

AllSides (2023). Allsides. `https://www.allsides.com`.

Alonso, M. A., Vilares, D., Gómez-Rodríguez, C., and Vilares, J. (2021). Sentiment analysis for fake news detection. *Electronics*, 10(11).

Amer, M. M. A. (2022). BBC and New York Times' coverage of the may 2021 Israeli onslaught on Gaza: A critical discourse analysis. *International Journal of English Linguistics*, 12(5):1–18.

Amos, A. and Haglund, M. (2000). From social taboo to "torch of freedom": The marketing of cigarettes to women. *Tobacco Control*, 9(1):3–8.

An, J., Kwak, H., and Ahn, Y.-Y. (2018). SemAxis: A lightweight framework to characterize domain-specific word semantics beyond sentiment. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2450–2461, Melbourne, Australia. Association for Computational Linguistics.

Ansar, A. and Khaled, A. F. M. (2023). In search of a Rohingya digital diaspora: Virtual togetherness, collective identities and political mobilisation. *Humanities and Social Sciences Communications*, 10(1):1–13.

Araque, O., Gatti, L., and Kalimeri, K. (2020). Moralstrength: Exploiting a moral lexicon and embedding similarity for moral foundations prediction. *Knowledge Based Systems*, 191(C).

Aslam, F., Awan, T. M., Syed, J. H., Kashif, A., and Parveen, M. (2020). Sentiments and emotions evoked by news headlines of coronavirus disease (Covid-19) outbreak. *Humanities and Social Sciences Communications*, 7(1):1–9.

Auxier, B. and Anderson, M. (2021). Social media use in 2021. Technical report, Pew Research Center.

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., Lee, J., Mann, M., Merhout, F., and 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.

Baly, R., Da San Martino, G., Glass, J., and Nakov, P. (2020). We can detect your bias: Predicting the political ideology of news articles. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4982–4991, Online. Association for Computational Linguistics.

Baumann, F., Lorenz-Spreen, P., Sokolov, I. M., and Starnini, M. (2020). Modeling echo chambers and polarization dynamics in social networks. *Physical Review Letters*, 124:048301:1–048301:6.

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

Bell, A. (1991). *The language of news media*. Blackwell Oxford.

Berger, J. and Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research*, 49(2):192–205.

Bläsi, B. (2009). Implementing peace journalism: The role of conflict stages. *Conflict & Communication*, 8(2).

Bode, L., Vraga, E. K., Borah, P., and Shah, D. V. (2014). A new space for political behavior: Political social networking and its democratic consequences. *Journal of Computer Mediated Communication*, 19:414–429.

Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V., and Kalai, A. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS, pages 4356—4364. Curran Associates Inc.

Boulianne, S. (2015). Social media use and participation: A meta-analysis of current research. *Information, Communication & Society*, 18(5):524–538.

Boxell, L., Gentzkow, M., and Shapiro, J. M. (2020). Cross-country trends in affective polarization. Working Paper 26669, National Bureau of Economic Research.

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

Branton, R. P. and Dunaway, J. (2009). Spatial proximity to the U.S.–Mexico border and newspaper coverage of immigration issues. *Political Research Quarterly*, 62(2):289–302.

Brounéus, K. (2011). *In-Depth Interviewing: The Process, Skill and Ethics of Interviews in Peace Research*, chapter 8, pages 130–145. Routledge.

Budak, C., Goel, S., and Rao, J. M. (2016). Fair and balanced? Quantifying media bias through crowdsourced content analysis. *Public Opinion Quarterly*, 80(S1):250–271.

Chen, T., Shi, J., Yang, J., Cong, G., and Li, G. (2020a). Modeling public opinion polarization in group behavior by integrating SIRS-based information diffusion process. *Complexity*, 2020:4791527:1–4791527:20.

Chen, W.-F., Al Khatib, K., Wachsmuth, H., and Stein, B. (2020b). Analyzing political bias and unfairness in news articles at different levels of granularity. In *Proceedings of the Fourth Workshop on Natural Language Processing and Computational Social Science*, pages 149–154, Online. Association for Computational Linguistics.

Chinn, S., Hart, P. S., and Soroka, S. (2020). Politicization and polarization in climate change news content, 1985-2017. *Science Communication*, 42(1):112–129.

Chuai, Y. and Zhao, J. (2022). Anger can make fake news viral online. *Frontiers in Physics*, 10:970174.

Cicchini, T., del Pozo, S. M., Tagliazucchi, E., and Balenzuela, P. (2022). News sharing on Twitter reveals emergent fragmentation of media agenda and persistent polarization. *EPJ Data Science*, 11(1):1–18.

Cinelli, M., Morales, G. D. F., Galeazzi, A., Quattrociocchi, W., and Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9):e2023301118.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Routledge.

Coscia, M. and Rossi, L. (2022). How minimizing conflicts could lead to polarization on social media: An agent-based model investigation. *PloS One*, 17(1):e0263184.

Cureton, E. E. (1956). Rank-biserial correlation. *Psychometrika*, 21(3):287–290.

DellaVigna, S. and Kaplan, E. (2007). The fox news effect: Media bias and voting. *The Quarterly Journal of Economics*, 122(3):1187–1234.

Dev, S., Li, T., Phillips, J. M., and Srikumar, V. (2021). OSCaR: Orthogonal subspace correction and rectification of biases in word embeddings. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 5034–5050, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Dev, S. and Phillips, J. (2019). Attenuating bias in word vectors. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 879–887. PMLR.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 17th Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, volume 1, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Dilliplane, S. (2011). All the news you want to hear: The impact of partisan news exposure on political participation. *Public Opinion Quarterly*, 75(2):287–316.

Dimock, M. and Wike, R. (2021). America is exceptional in its political divide. Technical report, The Pew Charitable Trust.

Dor, D. (2003). On newspaper headlines as relevance optimizers. *Journal of Pragmatics*, 35(5):695–721.

Doufesh, B. and Briel, H. (2021). Ethnocentrism in conflict news coverage: A multimodal framing analysis of the 2018 Gaza protests in the times of Israel and Al Jazeera. *International Journal of Communication*, 15:4230–4251.

Druckman, J. N. and Parkin, M. (2005). The impact of media bias: How editorial slant affects voters. *The Journal of Politics*, 67(4):1030–1049.

Dunaway, J. (2008). Markets, ownership, and the quality of campaign news coverage. *The Journal of Politics*, 70(4):1193–1202.

Dunaway, J. (2013). Media ownership and story tone in campaign news. *American Politics Research*, 41(1):24–53.

Elisa Shearer, E. G. (2019). Americans are wary of the role social media sites play in delivering the news. Technical report, Pew Research Center.

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4):51–58.

Evans, S. L., Jones, R., Alkan, E., Sichman, J. S., Haque, A., de Oliveira, F. B. S., and Mougouei, D. (2023). The emotional impact of COVID-19 news reporting: A longitudinal study using natural language processing. *Human Behavior and Emerging Technologies*, 2023(1):7283166.

Fahmy, S. and Eakin, B. (2014). High drama on the high seas: Peace versus war journalism framing of an Israeli/Palestinian-related incident. *International Communication Gazette*, 76(1):86–105.

Fan, L., White, M., Sharma, E., Su, R., Choubey, P. K., Huang, R., and Wang, L. (2019). In plain sight: Media bias through the lens of factual reporting. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6343–6349, Hong Kong, China. Association for Computational Linguistics.

Festinger, L. (1957). *A Theory of Cognitive Dissonance*, volume 2. Stanford University Press, Stanford, CA.

Field, A. (2013). *Discovering Statistics using IBM SPSS Statistics.* Sage, London, England.

Field, A. and Tsvetkov, Y. (2019). Entity-centric contextual affective analysis. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2550–2560, Florence, Italy. Association for Computational Linguistics.

Flaxman, S., Goel, S., and Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, 80:298–320.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5):378–382.

Freedman, J. L. and Sears, D. O. (1965). Selective exposure. *Advances in Experimental Social Psychology*, 2:57–97.

Frimer, J. A., Boghrati, R., Haidt, J., Graham, J., and Dehgani, M. (2019). Moral foundations dictionary for linguistic analyses 2.0. [Accessed 03-10-2025].

Fu, Z., Lam, W., Yu, Q., So, A. M.-C., Hu, S., Liu, Z., and Collier, N. (2023). Decoder-only or encoder-decoder? interpreting language model as a regularized encoder-decoder.

Gaitonde, J., Kleinberg, J., and Tardos, E. (2021). *Polarization in Geometric Opinion Dynamics*, page 499–519. Association for Computing Machinery, New York.

Galtung, J. (1986). On the role of the media in worldwide security and peace. *Peace and Communication*, pages 249–266.

Galtung, J. (1998). High road, low road: Charting the course for peace journalism. *Track Two: Constructive Approaches to Community and Political Conflict*, 7(4):7–10.

Galtung, J. and Ruge, M. H. (1965). The structure of foreign news: The presentation of the congo, cuba and cyprus crises in four norwegian newspapers. *Journal of Peace Research*, 2(1):64–90.

Garimella, K., Smith, T., Weiss, R., and West, R. (2021). Political polarization in online news consumption. *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM)*, 15(1):152–162.

Garrett, R. K., Gvirsman, S. D., Johnson, B. K., Tsfati, Y., Neo, R., and Dal, A. (2014). Implications of pro and counter-attitudinal information exposure for affective polarization. *Human Communication Research*, 40(3):309–332.

Garz, M., Sood, G., Stone, D. F., and Wallace, J. (2020). The supply of media slant across outlets and demand for slant within outlets: Evidence from US presidential campaign news. *European Journal of Political Economy*, 63.

Gelvin, J. L. (2021). *The Israel-Palestine Conflict: A History*. Cambridge University Press, Cambridge, 4 edition.

Gentzkow, M. and Shapiro, J. M. (2010). What drives media slant? Evidence from U.S. daily newspapers. *Econometrica*, 78(1):35–71.

Gerber, A. S., Green, D. P., and Shachar, R. (2003). Voting may be habit-forming: Evidence from a randomized field experiment. *American Journal of Political Science*, 47(3):540–550.

Gerber, A. S., Karlan, D., and Bergan, D. (2009). Does the media matter? A field experiment measuring the effect of newspapers on voting behavior and political opinions. *American Economic Journal: Applied Economics*, 1(2):35–52.

Ghasiya, P. and Okamura, K. (2021). Investigating COVID-19 news across four nations: A topic modeling and sentiment analysis approach. *IEEE Access*, 9:36645–36656.

Gouse, V., Valentin-Llopis, M., Perry, S., and Nyamwange, B. (2019). An investigation of the conceptualization of peace and war in peace journalism studies of media coverage of national and international conflicts. *Media, War & Conflict*, 12(4):435–449.

Graham, J., Haidt, J., and Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, 96(5):1029–1046.

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

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*.

Groseclose, T. and Milyo, J. (2005). A measure of media bias. *The Quarterly Journal of Economics*, 120(4):1191–1237.

Gu, W. J., Zhong, Y. H., Li, S. Z., Wei, C. S., Dong, L. T., yue Wang, Z., and Yan, C. (2024). Predicting stock prices with FinBERT-LSTM: Integrating news sentiment analysis. In *Proceedings of the 8th International Conference on Cloud and Big Data Computing*, pages 67–72, New York, NY, USA. Association for Computing Machinery.

Ha, L., Yang, Y., Ray, R., Matanji, F., Chen, P., Guo, K., and Lyu, N. (2020). How us and chinese media cover the us–china trade conflict: A case study of war and peace journalism practice and the foreign policy equilibrium hypothesis. *Negotiation and Conflict Management Research*.

Haidt, J. and Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, 20:98–116.

Haidt, J. and Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus*, 133(4):55–66.

Hajra Chaudhry, D. M. R. (2024). Ethical dilemmas in media coverage of Israel-Palestine conflict: Analysis of the New York Times. *Migration Letters*, 21(S11):72–83.

Hallin, D. C. (1989). *The Uncensored War: The Media and Vietnam*. University of California Press.

Hamborg, F. (2023). *Revealing Media Bias in News Articles: NLP Techniques for Automated Frame Analysis*. Springer Nature Switzerland, Cham, Switzerland.

Hamborg, F. and Donnay, K. (2021). NewsMTSC: A dataset for (multi-)target-dependent sentiment classification in political news articles. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 1663–1675, Online. Association for Computational Linguistics.

Hamborg, F., Donnay, K., and Gipp, B. (2021). Towards target-dependent sentiment classification in news articles. In *Diversity, Divergence, Dialogue*, pages 156–166, Cham. Springer.

Hamilton, J. T. (2003). *All the News That's Fit to Sell: How the Market Transforms Information into News.* Princeton University Press, Princeton.

Haque, A. and Singh, M. P. (2024). NewsSlant: Analyzing political news and its influence through a moral lens. *IEEE Transactions on Computational Social Systems*, 11(3):3329–333.

Hargittai, E., Gallo, J., and Kane, M. (2008). Cross-ideological discussions among conservative and liberal bloggers. *Public Choice*, 134(1-2):67–86.

Hart, W., Albarracín, D., Eagly, A. H., Brechan, I., Lindberg, M. J., and Merrill, L. (2009). Feeling validated versus being correct: A meta-analysis of selective exposure to information. *Psychological Bulletin*, 135(4):555–558.

Hassell, H. J. G., Miles, M. R., and Reuning, K. (2022). Does the ideology of the newsroom affect the provision of media slant? *Political Communication*, 39(2):184–201.

Heltzel, G. and Laurin, K. (2020). Polarization in america: two possible futures. *Current Opinion in Behavioral Sciences*, 34:179–184. Political Ideologies.

Hendrycks, D., Burns, C., Basart, S., Critch, A., Li, J., Song, D., and Steinhardt, J. (2021). Aligning AI with shared human values. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net.

Heni, A. and Chandra, O. (2022). The representation of Palestinian-Israeli conflict in online news articles: A critical discourse analysis. *Lensa: Kajian Kebahasaan, Kesusastraan, Dan Budaya*, 12:134–147.

Ho, D. E., Quinn, K. M., et al. (2008). Measuring explicit political positions of media. *Quarterly Journal of Political Science*, 3(4):353–377.

Hong, S. and Kim, S. H. (2016). Political polarization on Twitter: Implications for the use of social media in digital governments. *Government Information Quarterly*, 33(4):777–782.

Hoover, J., Portillo-Wightman, G., Yeh, L., Havaldar, S., Davani, A. M., Lin, Y., Kennedy, B., Atari, M., Kamel, Z., Mendlen, M., et al. (2020). Moral foundations Twitter corpus: A collection of 35k tweets annotated for moral sentiment. *Social Psychological and Personality Science*, 11(8):1057–1071.

Hopp, F. R., Fisher, J. T., Cornell, D., Huskey, R., and Weber, R. (2021). The extended moral foundations dictionary (emfd): Development and applications of a crowd-sourced approach to extracting moral intuitions from text. *Behavior Research Methods*, 53:232–246.

Hopp, F. R. and Weber, R. (2021). Reflections on extracting moral foundations from media content. *Communication Monographs*, 88(3):371–379.

Hu, Y., Hosseini, M., Skorupa Parolin, E., Osorio, J., Khan, L., Brandt, P., and D'Orazio, V. (2022). ConfliBERT: A pre-trained language model for political conflict and violence. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5469–5482, Seattle, United States. Association for Computational Linguistics.

Hussain, S., Siraj, S. A., and Mahmood, T. (2021). Evaluating war and peace potential of Pakistani news media: Advancing a peace journalism model. *Information Development*, 37(1):105–121.

Hązła, J., Jin, Y., Mossel, E., and Ramnarayan, G. (2019). A geometric model of opinion polarization. *arXiv*.

Ibrahim, F., Mustaffa, N., Ahmad, F., Kee, C. P., and Mahmud, W. A. W. (2013). Peace journalism: Implications of war and peace news amongst Malaysian audience. *Journal of Asian Pacific Communication*, 23(2):258–269.

Ifantidou, E. (2009). Newspaper headlines and relevance: Ad hoc concepts in ad hoc contexts. *Journal of Pragmatics*, 41(4):699–720.

Infield, T. (2020). Americans who get news mainly on social media are less knowledgeable and less engaged. Technical report, Pew Research Center.

Ishibashi, Y., Yokoi, S., Sudoh, K., and Nakamura, S. (2024). Subspace representations for soft set operations and sentence similarities. In Duh, K., Gomez, H., and Bethard, S., editors, *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 3512–3524, Mexico City, Mexico. Association for Computational Linguistics.

Ito, T. A., Larsen, J. T., Smith, N. K., and Cacioppo, J. T. (1998). Negative information weighs more heavily on the brain: The negativity bias in evaluative categorizations. *Journal of Personality and Social Psychology*, 75(4):887–900.

Iyengar, S. and Kinder, D. R. (1989). News that matters: Television & american opinion. *American Journal of Sociology*, 94(4):882–884.

Izzidien, A., Fitz, S., Romero, P., Loe, B. S., and Stillwell, D. (2022). Developing a sentence level fairness metric using word embeddings. *International Journal of Digital Humanities*, 5(2–3):95–130.

Jiang, C., Li, W., Bai, Q., and Zhang, M. (2017). Preference aware influence maximization. In *Multi-Agent and Complex Systems*, pages 153–164, Singapore. Springer.

Johnson, B. K., Neo, R. L., Heijnen, M. E., Smits, L., and van Veen, C. (2020). Issues, involvement, and influence: Effects of selective exposure and sharing on polarization and participation. *Computers in Human Behavior*, 104:106155.

Johnson, K. and Goldwasser, D. (2018). Classification of moral foundations in microblog political discourse. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 720–730, Melbourne, Australia. Association for Computational Linguistics.

Jones, J. M. (2018). Americans: Much misinformation, bias, inaccuracy in news. [Accessed 03-10-2025].

Kahn, K. F. and Kenney, P. J. (2002). The slant of the news: How editorial endorsements influence campaign coverage and citizens' views of candidates. *American Political Science Review*, 96(2):381–394.

Kalfeli, N., Frangonikolopoulos, C., and Gardikiotis, A. (2022). Expanding peace journalism: A new model for analyzing media representations of immigration. *Journalism*, 23(8):1789–1806.

Karamshuk, D., Lokot, T., Pryymak, O., and Sastry, N. (2016). Identifying partisan slant in news articles and Twitter during political crises. In *Social Informatics*, pages 257–272, Cham. Springer International Publishing.

Kelley, T. L. (1935). An unbiased correlation ratio measure. *Proceedings of the National Academy of Sciences of the United States of America*, 21(9):554–559.

Kempe, D., Kleinberg, J., and Tardos, E. (2003). Maximizing the spread of influence through a social network. In *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '03, page 137–146, New York. Association for Computing Machinery.

Kim, Y. (2015). Does disagreement mitigate polarization? How selective exposure and disagreement affect political polarization. *Journalism & Mass Communication Quarterly*, 92(4):915–937.

Kim, Y. (2019). How cross-cutting news exposure relates to candidate issue stance knowledge, political polarization, and participation: The moderating role of political sophistication. *International Journal of Public Opinion Research*, 31(4):626–648.

Kim, Y. and Chen, H.-T. (2016). Social media and online political participation: The mediating role of exposure to cross-cutting and like-minded perspectives. *Telematics and Informatics*, 33(2):320–330.

Kinder, D. R. and Sears, D. O. (1981). Prejudice and politics: Symbolic racism versus racial threats to the good life. *Journal of Personality and Social Psychology*, 40(3):414–431.

Klapper, J. T. (1960). *The Effects of Mass Communication*. Free Press, Glencoe.

Knobloch-Westerwick, S. (2014). *Choice and Preference in Media Use: Advances in Selective Exposure Theory and Research*. Routledge, New York.

Knobloch-Westerwick, S. and Johnson, B. K. (2014). Selective exposure for better or worse: Its mediating role for online news' impact on political participation. *Journal of Computer Mediated Communication*, 19:184–196.

Konnikova, M. (2014). How headlines change the way we think. *The New Yorker*, 17.

Kozitsin, I. V. and Chkhartishvili, A. G. (2020). Users' activity in online social networks and the formation of echo chambers. In *Proceedings of the 13th International Conference on Management of Large-Scale System Development (MLSD)*, pages 1–5. IEEE.

Kulkarni, R. V., Suryawanshi, S., Shirsath, A., Sonone, S., and Phapale, S. (2024). Online news: Event detection and community identification. In *ICT for Intelligent Systems*, pages 483–495, Singapore. Springer Nature.

Kull, S., Ramsay, C., and Lewis, E. (2003). Misperceptions, the media, and the Iraq war. *Political Science Quarterly*, 118(4):569–598.

Lacasse, K. and Forster, L. (2012). The war next door: Peace journalism in us local and distant newspapers' coverage of Mexico. *Media, War & Conflict*, 5(3):223–237.

Lazaridou, K., Krestel, R., and Naumann, F. (2017). Identifying media bias by analyzing reported speech. In *IEEE International Conference on Data Mining (ICDM)*, pages 943–948. IEEE.

Le, H., Shafiq, Z., and Srinivasan, P. (2017). Scalable news slant measurement using Twitter. *Proceedings of the International AAAI Conference on Web and Social Media*, 11(1):584–587.

Lee, S. T. (2010). Peace journalism: Principles and structural limitations in the news coverage of three conflicts. *Mass Communication and Society*, 13(4):361–384.

Lee, S. T. and Maslog, C. C. (2006). War or peace journalism? Asian newspaper coverage of conflicts. *Journal of Communication*, 55(2):311–329.

Lee, S. T., Maslog, C. C., and Kim, H. S. (2006). Asian conflicts and the Iraq war: A comparative framing analysis. *International Communication Gazette*, 68(5-6):499–518.

Lenhart, A. (2009). The democratization of online social networks. Technical report, Pew Research Center.

Leskovec, J. and Mcauley, J. (2012). Learning to discover social circles in ego networks. In *Advances in Neural Information Processing Systems*, volume 25, Lake Tahoe, Nevada. Curran Associates, Inc.

Levy, R. (2021). Social media, news consumption, and polarization: Evidence from a field experiment. *American Economic Review*, 111(3):831–70.

Li, W., Bai, Q., and Zhang, M. (2019). A multi-agent system for modelling preference-based complex influence diffusion in social networks. *The Computer Journal*, 62(3):430–447.

Lim, S., Jatowt, A., Färber, M., and Yoshikawa, M. (2020). Annotating and analyzing biased sentences in news articles using crowdsourcing. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 1478–1484, Marseille, France. European Language Resources Association.

Lim, S., Jatowt, A., and Yoshikawa, M. (2018). Understanding characteristics of biased sentences in news articles. In *CIKM Workshops*.

Lin, Z., Xie, J., and Li, Q. (2024). Multi-modal news event detection with external knowledge. *Information Processing & Management*, 61(3):103697.

Liu, Q., Zheng, Z., Zheng, J., Chen, Q., Liu, G., Chen, S., Chu, B., Zhu, H., Akinwunmi, B., Huang, J., Zhang, C. J. P., and Ming, W.-K. (2020). Health communication through news media during the early stage of the COVID-19 outbreak in China: Digital topic modeling approach. *Journal of Medical Internet Research*, 22(4):e19118.

Liu, S., Guo, L., Mays, K., Betke, M., and Wijaya, D. T. (2019a). Detecting frames in news headlines and its application to analyzing news framing trends surrounding U.S. gun violence. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, pages 504–514, Hong Kong, China. Association for Computational Linguistics.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. (2019b). RoBERTa: A robustly optimized BERT pretraining approach. *arXiv preprint*, abs/1907.11692.

Lord, C., Ross, L., and Lepper, M. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37:2098–2109.

Lynch, J. and McGoldrick, A. (2007). Peace journalism. In *Handbook of Peace and Conflict Studies*, pages 248–264. Routledge.

Mann, H. B. and Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 18(1):50–60.

Manzini, T., Yao Chong, L., Black, A. W., and Tsvetkov, Y. (2019). Black is to criminal as Caucasian is to police: Detecting and removing multiclass bias in word embeddings. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 615–621, Minneapolis, Minnesota. Association for Computational Linguistics.

Markus, G. B. (1982). Political attitudes during an election year: A report on the 1980 NES panel study. *The American Political Science Review*, 76(3):538–560.

Marozzo, F. and Bessi, A. (2017). Analyzing polarization of social media users and news sites during political campaigns. *Social Network Analysis and Mining*, 8.

Mathew, B., Sikdar, S., Lemmerich, F., and Strohmaier, M. (2020). The polar framework: Polar opposites enable interpretability of pre-trained word embeddings. In *'Proceedings of The Web Conference'*, page 1548–1558. Association for Computing Machinery.

McCombs, M. E. and Shaw, D. L. (1972). The agenda-setting function of mass media. *The Public Opinion Quarterly*, 36(2):176–187.

McGoldrick, A. and Lynch, J. (2000). Peace journalism – what is it? how to do it? [Accessed 03-10-2025].

McMahon, R. and Chow-White, P. A. (2011). News media encoding of racial reconciliation: Developing a peace journalism model for the analysis of 'cold' conflict. *Media, Culture & Society*, 33(7):989–1007.

Meffert, M., Chung, S., Joiner, A., Waks, L., and Garst, J. (2006). The effects of negativity and motivated information processing during a political campaign. *Journal of Communication*, 56:27–51.

Meta (2024). Huggingface. [accessed 02-05-2025].

Mokhberian, N., Abeliuk, A., Cummings, P., and Lerman, K. (2020). Moral framing and ideological bias of news. In *Social Informatics*, pages 206–219, Cham. Springer International Publishing.

Morales, A. J., Borondo, J., Losada, J. C., and Benito, R. M. (2015). Measuring political polarization: Twitter shows the two sides of Venezuela. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 25(3):033114:1–9.

Mutz, D. C. (2002a). The consequences of cross-cutting networks for political participation. *American Journal of Political Science*, 46(4):838–855.

Mutz, D. C. (2002b). Cross-cutting social networks: Testing democratic theory in practice. *American Political Science Review*, 96(1):111–126.

Nahon, K. (2015). *Where There Is Social Media There Is Politics*, chapter 3, pages 39–55. Routledge, New York.

Nardin, L. G., Balke-Visser, T., Ajmeri, N., Kalia, A. K., Sichman, J. S., and Singh, M. P. (2016). Classifying sanctions and designing a conceptual sanctioning process model for socio-technical systems. *The Knowledge Engineering Review (KER)*, 31(2):142–166.

Neumann, R. and Fahmy, S. (2016). Measuring journalistic peace/war performance: An exploratory study of crisis reporters' attitudes and perceptions. *International Communication Gazette*, 78(3):223–246.

Newell, E., Schang, A., Margolin, D., and Ruths, D. (2017). Assessing the verifiability of attributions in news text. In *Proceedings of the 8th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 754–763.

Newman, M. E. (2003). Mixing patterns in networks. *Physical Review E*, 67(2):026126.

Newsapi (2018). Rest api for searching and retrieving news articles. [accessed 05-24-2023].

Newspaper3k (2014). Newspaper3k: Article scraping & curation. [accessed 05-24-2023].

Nguyen, C. T. (2020). Echo chambers and epistemic bubbles. *Episteme*, 17(2):141–161.

Nugent, T., Petroni, F., Raman, N., Carstens, L., and Leidner, J. L. (2017). A comparison of classification models for natural disaster and critical event detection from news. In *IEEE International Conference on Big Data*, pages 3750–3759.

Okada, K. (2013). Is omega squared less biased? a comparison of three major effect size indices in one-way anova. *Behaviormetrika*, 40(2):129–147.

Oliveira, F. B., Haque, A., Mougouei, D., Evans, S., Sichman, J. S., and Singh, M. P. (2022). Investigating the emotional response to Covid-19 news on Twitter: A topic modeling and emotion classification approach. *IEEE Access*, 10:16883–16897.

Oliveira, F. B., Mougouei, D., Haque, A., Sichman, J. S., Dam, H. K., Evans, S., Ghose, A., and Singh, M. P. (2023). Beyond fear and anger: A global analysis of emotional response to Covid-19 news on Twitter. *Online Social Networks and Media*, 36:100253.

Ottosen, R. (2010). The war in afghanistan and peace journalism in practice. *Media, War & Conflict*, 3(3):261–278.

O'Connor, B., Balasubramanyan, R., Routledge, B., and Smith, N. (2010). From tweets to polls: Linking text sentiment to public opinion time series. *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM)*, 4(1):122–129.

Pariser, E. (2011). *The Filter Bubble: What the Internet Is Hiding From You.* Penguin UK.

Pew (2017). The partisan divide on political values grows even wider. Technical report, Pew Research Center.

Preprocessor, T. (2016). Tweet preprocessor: Elegant tweet preprocessing. [accessed 05-24-2023].

Price, V., Cappella, J. N., and Nir, L. (2002). Does disagreement contribute to more deliberative opinion? *Political Communication*, 19(1):95–112.

Project, M. I. (2018). Americans and the news media: What they do – and don't – understand about each other. `https://www.americanpressinstitute.org/wp-content/uploads/2018/06/Americans-and-the-News-Media-2018.pdf`. Accessed: 2021-03-17.

Pröllochs, N., Bär, D., and Feuerriegel, S. (2021a). Emotions explain differences in the diffusion of true vs. false social media rumors. *Scientific Reports*, 11(1):22721.

Pröllochs, N., Bär, D., and Feuerriegel, S. (2021b). Emotions in online rumor diffusion. *EPJ Data Science*, 10(1):51.

Puryear, C., Kubin, E., Schein, C., Bigman, Y. E., Ekstrom, P., and Gray, K. (2024). People believe political opponents accept blatant moral wrongs, fueling partisan divides. *PNAS Nexus*, 3(7):pgae244.

Qian, Y., Deng, X., Ye, Q., Ma, B., and Yuan, H. (2019). On detecting business event from the headlines and leads of massive online news articles. *Information Processing & Management*, 56(6):102086.

Qorib, M., Moon, G., and Ng, H. T. (2024). Are decoder-only language models better than encoder-only language models in understanding word meaning? In *Findings of the Association for Computational Linguistics*, pages 16339–16347, Bangkok, Thailand. Association for Computational Linguistics.

Rajpurkar, P., Jia, R., and Liang, P. (2018). Know what you don't know: Unanswerable questions for SQuAD. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 784–789, Melbourne, Australia. Association for Computational Linguistics.

Rajpurkar, P., Zhang, J., Lopyrev, K., and Liang, P. (2016). SQuAD: 100,000+ questions for machine comprehension of text. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 2383–2392, Austin, Texas. Association for Computational Linguistics.

Rea, L. M. and Parker, R. A. (2014). *Designing and Conducting Survey Research: A Comprehensive Guide*. John Wiley & Sons.

Redlawsk, D. P. (2002). Hot cognition or cool consideration? Testing the effects of motivated reasoning on political decision making. *The Journal of Politics*, 64(4):1021–1044.

Roberts, H., Bhargava, R., Valiukas, L., Jen, D., Malik, M. M., Bishop, C. S., Ndulue, E. B., Dave, A., Clark, J., Etling, B., Faris, R., Shah, A., Rubinovitz, J., Hope, A., D'Ignazio, C., Bermejo, F., Benkler, Y., and Zuckerman, E. (2021). Media cloud: Massive open source collection of global news on the open web. *Proceedings of the International AAAI Conference on Web and Social Media*, 15(1):1034–1045.

Robertson, C. E., Pröllochs, N., Schwarzenegger, K., Pärnamets, P., Van Bavel, J. J., and Feuerriegel, S. (2023). Negativity drives online news consumption. *Nature Human Behaviour*, 7(5):812–822.

Roy, S. and Goldwasser, D. (2021). Analysis of nuanced stances and sentiment towards entities of US politicians through the lens of moral foundation theory. In *Proceedings of the 9th International Workshop on Natural Language Processing for Social Media*, pages 1–13, Online. Association for Computational Linguistics.

Roy, S., Pacheco, M. L., and Goldwasser, D. (2021). Identifying morality frames in political tweets using relational learning. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 9939–9958, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Rozin, P. and Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, 5(4):296–320.

Schkade, D., Sunstein, C. R., and Hastie, R. (2007). What happened on deliberation day. *California Law Review*, 95:915.

Schramowski, P., Turan, C., Andersen, N., Rothkopf, C. A., and Kersting, K. (2022). Large pre-trained language models contain human-like biases of what is right and wrong to do. *Nature Machine Intelligence*, 4(3):258–268.

Shah, D. V., Watts, M. D., Domke, D., and Fan, D. P. (2002). News framing and cueing of issue regimes: Explaining Clinton's public approval in spite of scandal. *Public Opinion Quarterly*, 66(3):339–370.

Shapiro, A. H., Sudhof, M., and Wilson, D. J. (2022). Measuring news sentiment. *Journal of Econometrics*, 228(2):221–243.

Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611.

Sherif, M. and Hovland, C. (1961). *Social Judgment: Assimilation and Contrast Effects in Communication and Attitude Change*. Yale University Press, New Haven, CT.

Sheshadri, K., Shivade, C., and Singh, M. P. (2021a). Detecting framing changes in topical news. *IEEE Transactions on Computational Social Systems (TCSS)*, 8(3):780–791.

Sheshadri, K., Shivade, C., and Singh, M. P. (2021b). Detecting framing changes in topical news. *IEEE Transactions on Computational Social Systems*, 8(3):780–791.

Sheshadri, K. and Singh, M. P. (2019). The public and legislative impact of hyperconcentrated topic news. *Science Advances*, 5:1–8. Article number eaat8296.

Shi, E., Ding, L., Kong, L., and Jiang, B. (2024). Debiasing with sufficient projection: A general theoretical framework for vector representations. In Duh, K., Gomez, H., and Bethard, S., editors, *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 5960–5975, Mexico City. Association for Computational Linguistics.

Shinar, D. (2009). Can peace journalism make progress?: The coverage of the 2006 lebanon war in canadian and israeli media. *International Communication Gazette*, 71(6):451–471.

Siraj, S. A. (2008). War or peace journalism in elite us newspapers: Exploring news framing in pakistan-india conflict. *Strategic Studies*, 28(1):194–222.

Siraj, S. A. (2010). Framing war and peace journalism on the perspective of talibanisation in pakistan. *Media Asia*, 37(1):13–20.

statsmodel (2023). Granger causality test. [Accessed 03-10-2025].

Stein Tønnesson, M. Z. O. and Aung, N. L. (2022). Pretending to be states: The use of Facebook by armed groups in Myanmar. *Journal of Contemporary Asia*, 52(2):200–225.

Stroud, N. J. (2007). Media effects, selective exposure, and Fahrenheit 9/11. *Political Communication*, 24(4):415–432.

Stroud, N. J. (2010). Polarization and partisan selective exposure. *Journal of Communication*, 60(3):556–576.

Strovsky, D. and Schleifer, R. (2021). Playing with information: The Israeli-Palestinian conflict in the Russian press. *Middle East Policy*, 28(1):105–122.

Survey, A. G. F. (2018). Indicators of news media trust. [Accessed 03-10-2025].

Suwarno, S. and Sahayu, W. (2020). Palestine and Israel representation in the national and international news media: A critical discourse study. *Humaniora*, 32(3):217–225.

Taber, C. S. and Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50(3):755–769.

Tannenbaum, P. H. (1953). The effect of headlines on the interpretation of news stories. *Journalism Quarterly*, 30(2):189–197.

Tavoschi, L., Quattrone, F., D'Andrea, E., Ducange, P., Vabanesi, M., Marcelloni, F., and Lopalco, P. L. (2020). Twitter as a sentinel tool to monitor public opinion on vaccination: An opinion mining analysis from september 2016 to august 2017 in italy. *Human Vaccines & Immunotherapeutics*, 16(5):1062–1069. PMID: 32118519.

Tenenboim-Weinblatt, K., Hanitzsch, T., and Nagar, R. (2016). Beyond peace journalism: Reclassifying conflict narratives in the israeli news media. *Journal of Peace Research*, 53(2):151–165.

Tomczak, M. and Tomczak, E. (2014). The need to report effect size estimates revisited. an overview of some recommended measures of effect size. *Trends in Sport Sciences*, 21(1):19–25.

Tong, M., Wang, S., Cao, Y., Xu, B., Li, J., Hou, L., and Chua, T.-S. (2020). Image enhanced event detection in news articles. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(05):9040–9047.

Twitter (2021). Twitter developer api v2. [accessed 05-24-2023].

Tähtinen, T. (2024). When Facebook is the internet: The role of social media in ethnic conflict. *World Development*, 180:106633.

Ukka, I. (2023). International media role in Israel-Palestine conflict management: Analysis of media practitioner viewpoints. In *Explore Business, Technology Opportunities and Challenges After the Covid-19 Pandemic*, pages 1043–1055, Cham. Springer International Publishing.

Valentino, N. A. and Sears, D. O. (1998). Event-driven political communication and the preadult socialization of partisanship. *Political Behavior*, 20(2):127–154.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. u., and Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, Long Beach, CA. Curran Associates, Inc.

Vosoughi, S., Roy, D., and Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380):1146–1151.

Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., and Bowman, S. (2018a). GLUE: A multi-task benchmark and analysis platform for natural language understanding. In *Proceedings of the EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pages 353–355, Brussels, Belgium. Association for Computational Linguistics.

Wang, C., Koh, J. M., Cheong, K. H., and Xie, N.-G. (2019). Progressive information polarization in a complex-network entropic social dynamics model. *IEEE Access*, 7:35394–35404.

Wang, R. and Liu, W. (2021). Moral framing and information virality in social movements: A case study of #HongKongPoliceBrutality. *Communication Monographs*, 88(3):350–370.

Wang, Y., Ma, F., Jin, Z., Yuan, Y., Xun, G., Jha, K., Su, L., and Gao, J. (2018b). Eann: Event adversarial neural networks for multi-modal fake news detection. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, page 849–857. Association for Computing Machinery.

Weatherly, J. N., Petros, T. V., Christopherson, K. M., and Haugen, E. N. (2007). Perceptions of political bias in the headlines of two major news organizations. *Harvard International Journal of Press/Politics*, 12(2):91–104.

Westerwick, A., Johnson, B. K., and Knobloch-Westerwick, S. (2017). Confirmation biases in selective exposure to political online information: Source bias vs. content bias. *Communication Monographs*, 84(3):343–364.

Williams, A., Nangia, N., and Bowman, S. (2018). A broad-coverage challenge corpus for sentence understanding through inference. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1112–1122, New Orleans, Louisiana. Association for Computational Linguistics.

Wojcieszak, M., Bimber, B., Feldman, L., and Stroud, N. J. (2016). Partisan news and political participation: Exploring mediated relationships. *Political Communication*, 33(2):241–260.

Workneh, T. W. (2011). War journalism or peace journalism? a case study of us and british newspapers coverage of the somali conflict. *Global Media Journal: Mediterranean Edition*, 6(1).

Xi, R. and Singh, M. P. (2024a). The blame game: Understanding blame assignment in social media. *IEEE Transactions on Computational Social Systems (TCSS)*, 11(2):2267–2276.

Xi, R. and Singh, M. P. (2024b). Morality in the mundane: Categorizing moral reasoning in real-life social situations. In *Proceedings of the 18th International AAAI Conference on Web and Social Media (ICWSM)*, pages 1648–1660, Buffalo, New York. AAAI Press.

Yang, X., Chen, B.-C., Maity, M., and Ferrara, E. (2016). Social politics: Agenda setting and political communication on social media. In *Social Informatics*, pages 330–344, Cham. Springer International Publishing.

Yaqub, U., Chun, S. A., Atluri, V., and Vaidya, J. (2017). Analysis of political discourse on Twitter in the context of the 2016 US presidential elections. *Government Information Quarterly*, 34(4):613–626.

Yuan, G., Murukannaiah, P. K., Zhang, Z., and Singh, M. P. (2014). Exploiting sentiment homophily for link prediction. In *Proceedings of the 7th ACM Conference on Recommender Systems (RecSys)*, pages 17–24, Foster City, California. ACM.

Zaller, J. (1992a). *The Nature and Origins of Mass Opinion.* Cambridge University Press.

Zaller, J. R. (1992b). *The Nature and Origins of Mass Opinion.* Cambridge Studies in Public Opinion and Political Psychology. Cambridge University Press, Cambridge.

Zhang, Y., Ruan, X., Wang, H., Wang, H., and He, S. (2017). Twitter trends manipulation: A first look inside the security of Twitter trending. *IEEE Transactions on Information Forensics and Security*, 12(1):144–156.

Zhong, F., Li, P., and Xi, J. (2022). A survey on online political participation, social capital, and well-being in social media users—based on the second phase of the third (2019) TCS Taiwan communication survey database. *Frontiers in Psychology*, 12:730351–730351.

Zwillinger, D. and Kokoska, S. (1999). *CRC Standard Probability and Statistics Tables and Formulae.* CRC Press, Boca Raton.

APPENDICES

APPENDIX

A

REPRODUCING THE RESULTS

The codebase and data used in this dissertation are publicly available. The following are Github links containing all the resources required to reproduce the results for each chapter. Instructions for running the code are specified on GitHub.

Chapter 2 (NewsSlant: Analyzing Slants in Political News and User Engagement on Social Media)
Github link: https://github.com/ahaque2/NewsSlant

Chapter 3 (Moral Framing in War News: A Vector Subspace Approach)
Github link: https://github.com/ahaque2/Moral_Analysis_of_News

Chapter 4 (War or Peace Journalism: Analyzing Conflict News)
Github link: https://github.com/ahaque2/War_and_Peace_Journalism

Chapter 5 (Understanding Dynamics of Polarization via Multiagent Social Simulation)
Github link: https://github.com/ahaque2/MultiAgent_Social_Simulation

APPENDIX

B

NEWSSLANT

Keywords
Trump, Biden, Election, President, Democratic, Republican

Table B.1: Keywords used to identify news related to the 2020 US presidential elections.

Keywords
Donald, Trump, Joe, Biden, President, Republican, Democrate

Table B.2: Keywords removed during pre-processing before topic modeling.

Topic	**Keywords**
Covid-19	drug, outbreak, flu, infection, contagious, treatment, prescription, Covid, test, virus, ventilator, deaths, cases, pandemic, epidemic, corona, coronavirus, Covid19, patients, symptom
Covid-19 Lockdown	lockdown, shutdown, mask, distancing, masks
Covid-19 Vaccination	vaccine, vaccination, cure, moderna, pfizer, astrazeneca, j&j, dose
Covid-19 Educational_-impact	education, school, student, teacher, children, homeschool, schools, students, teachers
Economy	economy, industry, business, financial, finance, fiscal, economic, job, jobless, investing, investor, billion, gdp, debt, liquidity, inflation, stimulus, bill, stocks, market, employment, unemployment, checks, cheques, recession, bull, bullish, bear, bearish, dow, s&p, nasdaq, trade, trading, tax, loan, labor, buyback, selloff, wealth, wealthy, billionare, millionare'
Climate Change	earth, green, pollution, ozone, deforestation, greenhouse, wildfire, climate, warming, temperature, flood, drought, glacier, environment, environmental, carbon, emission, gas, fracking
Capitol Riots	capitol, riot, siege, rioter, mob
Elections	voter, absentee, ballot, fraud, mailin, stolen, voting, Election, black voters
Immigration	Immigration, immigrant, refugee, border, wall, migration, h1b, visa
Black Lives Matter	blm, floyd, police, brutality, defund, protest, protesters, officer, black lives matter, injustice, racism, racial, supremacist
Abortion	abortion, wade, roe
Supreme Court	supreme, court, coney, barret, packing, justice, judge
National Security	security, military, weapons, attack, defense, gun, shooting, pentagon
International News	international, country, global, china, chinese, beijing, shanghai, iran, irani, iranian, tehran, afghanistan, afghan, afghani, afghanistani, kabul, russia, russian, moscow, britain, british, brit, brexit, london, Korea, Korean, kim, venezuelan, venezuela, syrian, syriaworld, worldwide

Table B.3: List of seed words used for topic modeling news headlines and tweets.

Topic	SubTopics
Election	Elections, voting fraud, Election polls, black voters, georgia runoff
Covid-19	new cases, symptoms and precautions, vaccine, public advisory, drugs and treatment, testing, school reopenings, Covid-19 supplies
Economy	economic stimulus, taxes, markets

Table B.4: Topics formed by manually merging subtopics in news headlines. Subtopics are identified by BERTopic.

Topics	SubTopics
Election	Election, vote, Election fraud, electoral college, presidential debate, Election polls and opinions, biden inauguration, tulsa rally
Covid-19	new cases, Covid-19 response, face covering and mask, vaccine, school reopening, public advisory, Covid-19 treatment
Economy	economic stimulus, taxes
Conspiracy Theory	conspiracy theories, and fact-check

Table B.5: Topics formed by manually merging subtopics in news tweets. Subtopics are identified by BERTopic.

Pub	Headlines						Tweets					
	Trump			Biden			Trump			Biden		
	Pos	Neg	Neu	Pos	Neg	Neu	Pos	Neg	Neu	Pos	Neg	Neu
LEFT	0.110	0.568	0.322	0.275	0.218	0.507	0.129	0.598	0.274	0.329	0.230	0.441
CENTER	0.128	0.550	0.321	0.294	0.243	0.463	0.134	0.551	0.316	0.314	0.239	0.447
RIGHT	0.162	0.503	0.335	0.178	0.409	0.413	0.207	0.419	0.373	0.175	0.447	0.378

Table B.6: Mean sentiment scores of news headlines and tweets toward Trump and Biden for news publishers grouped based on political leaning.

Topic	Slant	Nonmoral	Care	Harm	Authority	Subversion	Fairness	Cheating	Loyalty	Betrayal	Sanctity	Degradation
Election	LEFT	−1	−13	−13	0	1	−1	9	1	−1	−6	−5
	BALANCED	−2	−9	−16	1	1	0	13	3	1	−9	−11
	RIGHT	1	−9	−16	2	−1	−1	5	4	−4	−5	−8
Covid-19	LEFT	5	18	19	−7	−8	−9	−10	−7	−9	−2	5
	BALANCED	4	16	14	−11	−8	−12	−15	−6	−12	−5	32
	RIGHT	8	19	13	−7	−12	−8	−16	−9	−15	−8	6
BLM	LEFT	−5	10	47	−2	−3	20	−4	2	22	0	−3
	BALANCED	−15	6	80	3	6	18	9	−3	35	0	6
	RIGHT	3	16	25	−7	−12	−1	−14	2	−5	85	19
Supreme Court	LEFT	−5	−5	−18	18	4	58	7	1	−6	14	−3
	BALANCED	−6	−18	−11	19	16	11	−2	−3	−5	9	24
	RIGHT	−7	−10	−19	23	10	34	14	−1	−1	2	−11
Economy	LEFT	−5	−5	−24	1	8	−6	23	1	1	−17	−17
	BALANCED	−10	−2	−13	7	15	−3	30	−4	5	−24	−14
	RIGHT	−13	22	−5	14	18	1	24	1	14	−12	−15
Conspiracy Theory	LEFT	−3	−19	−12	−10	−1	−3	21	−10	−3	1	1
	BALANCED	14	−12	−30	−24	−24	−6	6	−11	−19	−1	−24
	RIGHT	5	−5	−9	20	0	−14	−14	1	−9	−5	−4
Capitol Riots	LEFT	−13	30	40	9	18	0	−10	6	46	−3	9
	BALANCED	−15	−3	57	5	22	5	0	−3	56	3	−1
	RIGHT	−13	−3	65	0	16	12	−5	0	55	−10	2
Impeachment	LEFT	−7	−12	−18	24	23	4	−7	0	6	−2	15
	BALANCED	−11	13	−3	40	22	16	−6	26	17	17	5
	RIGHT	−12	−1	14	27	29	0	−9	33	26	−6	4
Healthcare	LEFT	−5	14	4	8	8	7	4	−2	−2	−6	−1
	BALANCED	−2	−9	44	−5	−5	−5	0	−4	−5	42	−2
	RIGHT	4	−1	−2	1	0	12	−2	−2	−3	−26	−30
Immigration	LEFT	3	32	34	−9	−12	6	−16	−1	−8	10	16
	BALANCED	−7	27	27	7	18	9	−12	60	23	33	−34
	RIGHT	−7	76	21	20	16	10	−13	25	21	−17	−10

Table B.7: Shift in the mean moral foundation scores for each topic from the overall mean for a given news publisher group. Values are in percent (%).

APPENDIX

C

MORAL FRAMING IN WAR NEWS

	Keywords
Include	Gaza, Hamas, Israel, IDF, Palestine, Al-Aqsa, Middle East, Ceasefire, Refugees, Airstrikes, West Bank, Rafah, Mossad, Netanyahu, Haniyeh, Jihad, militant, terrorist
Exclude	Ukraine, Russia, Kyiv, NATO, Crimea, Donetsk, Luhansk, Kherson, Zelensky, Putin, Wagner, Donbas, Europe

Table C.1: Keywords-based query used to identify relevant news headlines.

Groups	Entities
Palestine	Palestinians, Palestinian, Palestino, Rafah, West Bank, Palestine, Palestina, Pro-Palestinian, Pro-Palestine, Middle East, Jenin, Khan Younis, Gazans, Jabalia, Al-Shifa, Arab, South Africa, Rashida, Tlaib, Nakba, Kamal, Adwan, Nuseirat, AOC, Qatar, Ireland, China, France, Sudan
Hamas	Hamas, Hamás, pro-Hamas, Hezbollah, Yahya, Sinwar, Ismail, Haniyeh, Muslim, ISIS, Hassan, Nasrallah, Houthi, Iran, Iranian, Tehran, Lebanon, Lebanese, Beirut, Syria, Houthi, Houthis, Yemen, Khamenei
Israel	Israel, Israeli, Israelis, Israelí, İsrail, Israël, Isra'ila, Pro-Israel, Antisemitism, US, Biden, Netanyahu, Anti-Israel, IDF, Jewish, Trump, Blinken, Netanyahu, American, Americans, America, White House, Tel Aviv, Kamala, Harris, Jews, Jerusalem, GOP, Republicans, Republican, Democrats, FBI, Holocaust, DNC, Pentagon, DeSantis, DC, Washington
Unrelated	UN, WHO, UN Security Council, UK, ICJ, Egypt, BBC, UNRWA, ICC, London, Columbia, Harvard, British, EU, UCLA, Germany, CNN, Aljazeera, Fox News, Starmer, Jordan, Turkey, Scotland, Egyptian, NYPD, New Orleans, Texas, New York, Chicago, California, NYC, Congress

Table C.2: Entity Groups.

	LLM Prompt
Prompt 1	Identify whether a given headline reports on or relates to the ongoing conflict in Gaza. The conflict began on October 7, 2023, when Hamas launched a surprise attack on Israel, prompting retaliation from the Israel Defense Forces (IDF), who entered the Palestinian territory. A headline should be classified as related if it mentions any conflicted entity such as Israeli, Gaza, Palestinians, Hamas, Middle East, West Bank, Rafah, Tel Aviv, Netanyahu, Haniyeh, etc. A headline should also be classified as related if it provides updates on the conflict, states actions or statements from involved parties, states comments from external entities, direct or indirect consequences of the conflict, describes the historical context of the conflict, or its broader impact on uninvolved parties. Output a label '1' for headlines that are relevant or related and '0' if it is irrelevant or unrelated. Only return the labels and nothing else.
Prompt 2	Identify whether a given headline mentions any of the following entities: Gaza, Palestine, Palestinians, Israel, Israelis, Hamas, Rafah, West Bank, and any entity indirectly associated with these entities. For instance, IDF (Israel Defense Force) indirectly refers to Israel. Similarly, Israeli PM Netanyahu or the Defense Minister indirectly refers to Israel. Similarly, references like jihadists, terrorists, and militants indirectly refer to Hamas. Similarly, leaders of Hamas like Haniyeh refer indirectly to Hamas. Output a label '1' for headlines that mention an entity described above directly or indirectly, or if any headline can be associated with these entities, and '0' if the headline doesn't mention any above entities. Only return the label and nothing else.

Table C.3: Prompts to identify relevant news.

Moral Foundation	**High (*h*)**	**Low (*l*)**	**Number of word pairs (*k*)**
Care	275	475	175
Fairness	350	475	325
Loyalty	300	375	175
Authority	350	475	275
Sanctity	275	375	225

Table C.4: Optimal Parameters to Build the Moral Subspace Identified by Grid Search. For each parameter, we search between 25 and 500 with an increment of 25.

APPENDIX

D

WAR AND PEACE JOURNALISM ANNOTATION INSTRUCTIONS

D.1 Annotation Setup

Task. Read news headlines related to a conflict and annotate for war and peace journalism(wiki) and identify victim and villain

D.1.1 War Journalism

- Reinforces divisions between conflicted entities (e.g., "us vs. them" mentality)
- Focuses on visible effects of the conflict, such as military actions, violence, and destruction
- Sensationalizes news by using emotion-arousing language and promotes stereotypes
- Presents information from elite sources
- Brings out problems and differences that exist between the conflicted entities and focuses on here and now instead of providing broader historical context and potential solutions

For example,

Hamas Bears the Blame for Every Death in This War
Use of demonizing and charged-up language

DeSantis says US shouldn't take in Palestinian refugees from Gaza because they're 'all anti-semitic'
Unwarranted generalizations and stereotypes

D.1.2 Peace Journalism

- Humanizes all parties involved in the conflict and portrays the conflict as a problem instead of a conflicted entity.
- Focuses on invisible (indirect) effects of conflict, such as humanitarian and economic impact, psychological impact and trauma, population displacement, and cultural erosion.
- Presents news objectively, avoiding emotion-laden opinionated language
- Provides a voice to common people instead of just elites.
- Discusses the broader historical context of the conflict and its causes, highlights efforts to negotiate, discusses potential solutions, and emphasizes conflict resolution through diplomatic means such as dialogue

For example,

Palestinian mother fears for her children as she wonders about the future after evacuating Gaza City
Shows the trauma of war and its invisible effects

US tells UN Security Council Gaza ceasefire 'in sight' ahead of Cairo talks
Shows hope for a diplomatic solution to the conflict

D.1.3 Annotation Context

We focus on the news about the ongoing war in Gaza that started in October 2023. The ongoing conflict in Gaza is primarily between Israel and Hamas, a militant group that controls the Gaza Strip. The current conflict escalated dramatically after Hamas launched a major attack on October 7, 2023, killing around 1,200 Israelis, the majority of whom were civilians, and taking more than 240 hostages. In response, Israel initiated a large-scale military operation in Gaza, including aerial bombardments and a ground offensive (wiki).

Historical Context and Causes

The roots of the conflict are deep and complex, stemming from decades of tension between Israelis and Palestinians over territorial, political, and nationalistic disputes. Since the 2007 takeover of Gaza by Hamas following a brief civil war with the Palestinian Authority, Gaza has been under a blockade by both Israel and Egypt, which has severely limited the movement of people and goods. Hamas's militant activities, coupled with the blockade, have led to repeated cycles of violence (wiki).

Entities Involved in the Conflicted

Israel. The Israeli government and its IDF (Israel Defense Forces) are engaged in intense operations against Hamas in Gaza (Palestine), citing the need to neutralize the militant threat and prevent further attacks on Israeli territory.
Hamas. The militant group has been launching rockets into Israeli cities and is responsible for the initial attack that escalated the current conflict.
Palestinians. The most affected group, with widespread casualties and a dire humanitarian situation. Nearly 90% of Gaza's population is now facing food insecurity, and critical infrastructure, including hospitals and water treatment facilities, has been heavily damaged.

Humanitarian Impact

The conflict has had devastating effects on Gaza's civilian population. Thousands have been killed, and more than 1.7 million people have been displaced within the territory. Critical shortages of food, water, and medical supplies are contributing to a severe humanitarian crisis. The blockade and the ongoing military operations have caused widespread destruction of roads, residential buildings, and hospitals.

The long-standing blockade and intermittent conflicts have left Gaza in a state of economic and infrastructural collapse, and the civilian toll has been immense, with no end in sight to the violence and suffering on both sides.

Role of News Media

Media coverage of the Israel-Hamas conflict has varied widely, often influenced by different political and social contexts, resulting in distinct narratives. Western media tends to frame the conflict through a lens that aligns with Israel. They frequently describe Hamas as a "terrorist group" and focus on the impact of the conflict on Israeli civilians, highlighting the tragic loss of life and destruction caused by Hamas's initial Oct 7th attack. This portrayal often frames Israel's actions as self-defense, casting the Israeli military's subsequent operations as justified responses to terrorism. In contrast, Middle Eastern news media typically highlight the humanitarian impact of the conflict on Palestinians. Its reports emphasize civilian casualties in Gaza, the destruction of homes, and the blockade's effect on the region's population.

Media coverage of the conflict acts as a powerful tool that can shape public opinion on an ongoing conflict. This is especially applicable when the conflict is unfolding far away. The public often perceives war through the selective lens of what the media chooses to show and emphasize. The disparity in the media narrative and the portrayal of the conflicted entities can have real-world consequences. For example, U.S. audiences may rally behind the idea of increased military aid to Israel, whereas Middle Eastern audiences are more likely to push for international intervention to protect Palestinian civilians. The contrasting coverage also leads to misunderstandings and polarized viewpoints, making it harder to achieve a consensus on diplomatic solutions. When the news narrative about a conflict is skewed or selective, it creates an "information war," where public support or opposition to the conflict hinges less on the realities on the ground and more on how the war is presented in the news.

D.2 Annotation Task

We conduct an entity-centric annotation of the news headline on the ongoing Gaza conflict and evaluate the portrayal of the three conflicted entities, Israel, Hamas, and Palestine. We list a few keywords for each conflicted entity that can be considered representative of the entity. The list is not comprehensive, and you are allowed to add more entities that apply. The entity grouping is as follows.

Israel. Israel, Israeli, IDF, Netanyahu, Jew, Jewish, Hebrew, Zionist, Settler

Palestine. Gaza, Palestine, Palestinian

Hamas. Hamas, Haniyeh, Islamist, Terrorist, Jihadist, Militant

All entities other than the conflicted entities are nonconflicted entities. We evaluate news headlines related to the conflict and broadly fit into one of the following.

- Describe the actions of some conflicted entity (including comments, plans, and speculations) and their impact
- Describe the impact of some event on the conflicted entities
- Describe actions (including comments, plans, and speculations) by a nonconflicted entity that relate to the conflict (or the conflicted entities)

- Describe the impact of the conflict on nonconflicted entities

Each headline has to be annotated with five labels as described below.

D.2.1 Label 1: War and Peace Frames

For each headline, we identify the framing of the news. Framing is the way news is presented to evoke a particular interpretation or reaction from the audience. It shapes how information is structured, what details are emphasized, and how audiences perceive the significance of an event or issue. We identify the following four frames in news headlines that can help us differentiate between war and peace journalism.

Label 1a: Emotiveness of the headline

Identify if the news headline is *sensationalized* or *objective*.

Sensationalization(S). Refers to presenting information in an exaggerated, dramatic, or provocative manner to attract more attention. News publishers are often incentivized to sensationalize news as it attracts more views and generates more revenue. Sensationalization can take various forms, such as using emotionally charged language and focusing disproportionately on negative or shocking aspects or labeling entities involved in the conflict as good or bad (i.e. using demonizing or victimizing language). Additionally, news publishers often use sensational quotes to convey their agenda. Although one can argue they bear no responsibility for directly quoting someone, often such tactics are used to create sensational content and a war of words between entities involved directly or indirectly. Hence, we consider such headlines that use emotive quotes sensational as well. Following, we discuss some examples of how news headlines can be sensationalized.

Articulate a headline as a question. For example,

Is Israel criminalizing Palestinian thoughts too, amid Gaza war?

However, some headlines articulated as a question may not be sensationalized.

What next for Gaza and the talks to end the war?
Is US support for Israel damaging Washington's global standing?

Quoting sensational comments relating to the conflict. For example,

'We'll fight with fingernails' says Israeli PM after US warning
UN chief Guterres urges ceasefire as Gaza becomes 'graveyard for children'

Use of words or style that adds sensationalization. For example, *WATCH: White House shuts down question on authenticity of photos showing dead Israeli children*

Anti-Israel organizers at George Washington University issue new demand as campus takeover reaches 13th day

Objective(O). Headlines that avoid using emotive language and present information neutrally (i.e., without taking sides). Objective language is based on facts, presents factual information without exaggeration, and does not contain demonizing or victimizing language. Some examples of objective headlines are as follows.

Anti-government protesters demand new elections in Israel amid Gaza war

Hamas terrorists open fire at bus stop near Jerusalem, leaving 3 dead, 11 injured

When annotating this class focus on how the news is framed rather than the content of the news itself. Ask yourself: Can the same information be presented in a more neutral and objective way? If so, the headline may be sensationalized. If not, it is objective. The key point here is not to focus on the emotion the headline evokes, which is influenced by the news event itself, but rather how the information is framed—particularly the language used to present it. Does the language make the headline more emotional or dramatic, thus sensationalizing it? Or does the language present the information in a way that avoids appealing to emotions, aiming for clarity and neutrality?

For instance, consider the following examples where the same headline is presented as an objective headline followed by two sensationalized versions.

Israeli Airstrikes on Gaza Result in 35 Deaths (Objective)
Gaza Under Siege: Israeli Airstrikes Wreak Havoc, Death Toll Mounts! (Sensationalized)
Bloodshed in Gaza: Israeli Bombings Claim 35 Innocent Lives! (Sensationalized)

Humanitarian Aid Delivered to Gaza Amid Calls for Temporary Ceasefire (Objective)
Desperate Relief: Much-Needed Humanitarian Aid Floods into Gaza as Ceasefire Hopes Flicker! (Sensationalized)
Life-Saving Aid Finally Reaches Desperate Gaza as Ceasefire Hangs in the Balance (Sensationalized)

Headlines that seem to be picking a side should be annotated as Sensationalized as they promote the us-vs-them mentality in a conflict. Focus on the framing rather than the news event itself when determining this.

Label 1b: Effect Frame

Any headline describing the effects of war, such as casualties, destruction, economic sanctions, population displacement, etc. This includes speculations about the future, an event caused by the conflict, or a consequence of the conflict. The effect frame captures what the headline describes. This can be classified in two types, the visible effects of war, and the invisible effects of war.

Visible Effects of War (VE). Headlines that describe (or relate to) the visible effects of war, i.e., direct effects of war, such as death and destruction of property. This includes headlines related to acts of aggression, such as ground invasion, airstrikes, raids, etc. For example,

Overnight Israeli air raids kill 55 people in Gaza, authorities say

Israeli music festival survivor describes horror of Hamas-led attack that left 260 dead

Israeli forces raid Jenin, surround Ibn Sina hospital in occupied West Bank

'Bombing everywhere': As Israel renews war, our eight-year-old has questions

Invisible Effects of War (IE). Headlines that describe (or relate to) indirect effects of war that are often not as visible as the violence and destruction caused directly by the war. This includes psychological trauma, social fragmentation, population displacement, cultural erosion, food and health crises, etc. This also includes economic and other sanctions that may be put on the conflicted entities and events such as protests (related to the conflict). For example,

Is Israel criminalising Palestinian thoughts too, amid Gaza war?

Minimal' fuel to be allowed into Gaza after UN warns of starvation risk

Palestinian mother fears for her children as she wonders about the future after evacuating Gaza City

Israeli leaders condemn expected US sanctions, Netanyahu vows to fight it with all his might

Anti-government protesters demand new elections in Israel amid Gaza war

None. If neither (or both) frame applies or if the frame is unclear. For example,

Son of Hamas founder insists 'no such thing' as Israel, Hamas cease-fire until ayatollah is gone

Israel's war on Hamas: List of key events, day 83

Israel-Gaza war: Does the West have the power to stop Israel?

Historian Rick Richman On Israel's Founding As An Act of Passion: Just Politics

To differentiate between visible and invisible effects, consider whether the news describes a direct, immediately observable consequence of the ongoing war (focusing on the current events) or an indirect effect that may not be immediately apparent because it takes time to unfold or may not be visible in the same way. For example, emotions like fear and anger triggered by the war are less observable compared to physical destruction and casualties. Visible effects (VE) of war include military actions, casualties, and physical destruction, as these are immediate and tangible outcomes of war. On the other hand, invisible effects (IE) such as psychological trauma, social fragmentation, generational poverty, and crises in food and health often take time to manifest and are not as directly visible. Additionally, responses or reactions to the conflict, such as protests, economic sanctions, or other measures by entities not directly involved, should also be considered invisible effects.

Label 1c: Source Frame

The source frame identifies who the headline is providing a voice to, the elite or the common people. The source frame captures the source attribution (i.e., the source of the headline) and who the headline is about (i.e., the subject and object). Use the order of preference as source > subject > object in case of conflict. For instance, if the source indicates a different frame than the subject and object, for example, if the source is elite and the subject and object are common people, we label it as elite.

Elite sources and voices (ES). Headlines that provide a platform for the authorities and elite voices (and stories) rather than common people. This includes comments made by elite sources and information that is credited to some elite source. Consider all entities referenced with a name as elite. For example,

Israel's Netanyahu rips Ireland, Spain and Norway recognizing Palestinian statehood: 'Reward for terrorism'

Overnight Israeli air raids kill 55 people in Gaza, authorities say

Son of Hamas founder insists 'no such thing' as Israel, Hamas cease-fire until ayatollah is gone

Common people as sources (PS). Headlines that provide a voice to common people and describe peoples' stories and opinions. This includes headlines describing personal stories, common people's acts, and comments. For example,

'Bombing everywhere': As Israel renews war, our eight-year-old has questions

Muslim immigrant running to unseat Ilhan Omar, speaks out from Israel: 'We are not like her'

Israeli music festival attendee details horror of Hamas attack

Palestinian mother fears for her children as she wonders about the future after evacuating Gaza City

None. Headlines that do not have any source attributions, or if neither frames apply, or if the frame is unclear. This includes situations where the headline is ambiguous, or you lack the context to determine the source. For example,

Who is Israel targeting in its attacks on the West Bank?

Israel ramps up airstrikes on the Gaza Strip

Israel's war on Gaza: The view from China

The source frame focuses on determining whether a headline is giving a platform to elite sources (ES), amplifying the voice of the common people (PS), or simply reporting a news event (None). If a news publisher reports an event without citing a source, they are the source—these headlines should be annotated as None (this includes when they cite another news publisher as a source). However, when the publisher quotes someone or cites some authority or organization as the source, they shift responsibility away from the news publisher. We label such headlines as ES. If the news covers stories about individual struggles or collective movements, like protests, such headlines should be annotated as PS.

Note: We do not consider news sources as elite sources, i.e., having a source attribution to a news publisher doesn't make the headline ES. So, if a news is attributed to any other news publisher, such headlines should be annotated as None (unless it is a people story in which case it should be annotated as PS).

The source frame is not about the entity reported on. For instance, a headline such as

The people are protesting against President Biden

Should be annotated as PS due to the news report being about people (hence providing people a voice), rather than ES due to the presence of Biden. However, consider headlines that quote an elite, such as,

Israel's Netanyahu rips Ireland, Spain and Norway recognizing Palestinian statehood: 'Reward for terrorism'

The above headline provides a platform for the elite (Netanyahu), hence ES. Although one can argue a comment by the leader of a conflicted entity should be considered as news, we assume all comments by a named entity (including individuals, and organizations) as ES to simplify the task.

Label 1d: Context Frame

War Context Frame (WC). Headlines that focus on the latest developments in the ongoing conflict. These headlines describe the actions or comments of the conflicted entities or the effects of some event on them. For example,

Israeli forces arrest 28 Palestinians in raids in occupied West Bank

Overnight Israeli air raids kill 55 people in Gaza, authorities say

Israel takes control of Rafah crossing, Gaza's lifeline: What's going on?

Broader Context Frame (BC). Headlines that provide a broader context of the ongoing conflict, including historical contexts, or highlight the events that triggered the war. Headlines that describe the actions (or comments) of nonconflicted entities on the conflict. For example,

Rep. Michael McCaul says Egyptians warned Israel about attack

US expected to announce more support for Israel

South Africa files case at ICJ accusing Israel of 'genocidal acts' in Gaza

'Outraged': Brazilian Muslims face growing Islamophobia over Gaza war

None. When the context doesn't correspond to either of the context frames above or when the context is unclear or ambiguous. For example,

'Fire and forget' in Gaza

Gaza and the dilemmas of genocide scholars

Narrating Gaza: Pain in Arabic, information in English

Context frame captures the context of the headline. To identify the appropriate context frame that applies to a headline, ask yourself if the headline provides an update on the latest events in the ongoing war, i.e., war context frame (WC), or focuses on historical context and how the war impacts indirectly related entities and nonconflicted entities, i.e., broader context frame (BC). Whereas WC focuses on now and here, i.e., the current updates about the war, including actions and response (comments) of the conflicted entities, BC focuses on broader context (BC), such as highlighting reasons for the war, historical context, and impact of war on indirectly involved entities or nonconflicted entities (this includes responses to the conflict, such as comments and protests).

Some headlines from that are examples of BC (broader context) are as follows.

The horrifying reason why there are no Christmas lights in Bethlehem of all places UN expert accuses Israel of 'genocide' in Gaza US says Gaza pier being repaired as aid groups urge more *consistent routes Arizona State places police chief on leave amid complaints of alleged actions during anti-Israel protest*

D.2.2 Label 2: Role-labeling of conflicted entities

Identify which conflicted entity is portrayed as a villain and which as a victim. We define a villain as an entity that is shown in a bad light (i.e., negatively). In the context of war news, we consider the aggressor (the one attacking) as a villain. We define a victim as a conflicted entity that suffers harm, injustice, and adversity. In the context of war news, a victim refers to the conflicted entity that suffers physical, psychological, social, or economic harm as a direct or indirect consequence of conflict.

For each headline, provide annotation for the villain and the victim using the following labels.

Israel. Israel is portrayed as a villain or victim

Hamas. Hamas is portrayed as a villain or victim

Palestine. Palestine is portrayed as a villain or victim

None. No entity is portrayed as a villain or victim or you can't determine (this could be due to ambiguity in the headline or due to a lack of context)

Here are some examples.

Anti-government protesters demand new elections in Israel amid Gaza war

Israel is portrayed as a villain (protestors express disappointment and demand change)

Overnight Israeli air raids kill 55 people in Gaza, authorities say

Israel is portrayed as a villain (aggressor), and Palestinians (people in Gaza) as victims.

New Hampshire police arrest 3 pro-Palestinian protesters after building damaged at Israeli-owned company

Palestine is portrayed as a villain (through indirect referencing) and Israel (through indirect referencing) as a victim.

Israel-Gaza war: Does the West have the power to stop Israel?

No entity is portrayed as either a villain or a victim.

When determining the victim and villain, make sure to not 'over' infer. That is, annotate a villain and a victim if the headline portrays them so (i.e., mentions something to infer both). Having one doesn't ensure the presence of the other. For example, in H1, Israel is portrayed as a villain because people are protesting against the government (which is waging the war in Gaza) and demanding an election. One could argue this makes people living in Gaza (i.e., Palestinians) victims. However, we avoid such inferences and only annotate based on what is presented in the headline, using minimal external context to interpret the headlines.

Headlines that describe an act of aggression but explicitly provide a justification for the aggression should be considered as not depicting any villain. For example,

Israel responds to Hezbollah attack from Lebanon deploying 'tens of thousands' to northern border.

Explanation: The headline frames the action of Israel as a response to the attack by Lebanon, justifying the attack. Hence, Israel isn't a villain in the above headline.

For headlines that mention an act of aggression but do not specify the victims (or don't expand on the sufferings of the victims), we annotate these as having a villain but no victim. For instance, for a headline such as

UN expert accuses Israel of 'genocide' in Gaza

Explanation: In the above headline, Israel is portrayed as a villain, however, no victim is described. One could argue that the victim is Palestine since the headline is explicit about where this genocide is happening, i.e., 'in' Gaza. However, the focus of the headline is clearly to report on the aggressions by Israel and not portray the victims in Gaza (these aggressions are happening in Gaza, adding context). However, for a headline such as

'It's not complex, it's genocide': Former top UN official on Gaza

You should infer Palestine as a victim (and no villain). This is because the focus is to point the sufferings of the people living in Gaza (Palestinians) due to the genocide.

Headlines can have the same entity as the victim and the villain. For example,

Prisoner exchange? Israeli captive families demand answers from Netanyahu

The villain and the victim are both Israel. This is because an Israeli is a captive (hence victim), and the demand for an answer is made to Netanyahu, who we identify as a representative entity for Israeli, hence the villain is also Israel.

APPENDIX

E

DYNAMICS OF POLARIZATION

Metric	Description
Negative Satisfied	Agents with user satisfaction less than zero
Zero Satisfied	Agents with user satisfaction equal to zero
Positive Satisfied	Agents with user satisfaction greater than zero
Low Activity	Agents with user activity lower than or equal to 0.75
Medium Activity	Agents with user activity between [0.75, 0.90]
High Activity	Agents with user activity greater than or equal to 0.90
Low Polarized	Agents with POV in $[-0.1, 0.1]$
High Polarized	Agents with POV greater than 0.1 or lower than -0.1

Table E.1: Secondary metrics to compare initial and final user distribution based on agent's state.

Notation	Description
c_1	A constant (scale factor) to scale up smaller values. We use the value of 10.
c_2	A constant (scale factor) to scale down the larger values. We use the value of 0.1.
a_x	Agent x
p_k	k^{th} post shared on the social network
uS(a_x, i, p_k)	Stance of a_x toward issue i after p_k is shared
pS(p_k, i)	Stance of p_k toward issue i
uA(a_x, p_k)	Activity score for a_x after p_k is shared
sPref(a_x, p_k)	Sharing preference of a_x after p_k is shared
sP(a_x, p_k)	Probability of a_x to share p_k
Sanc(a_y, p_k, a_x)	Sanction a_y provides on receiving p_k from a_x
$\delta S(a_x, a_y, i, p_k)$	difference in stance between the spreader (a_x) and the receiver (a_y) on issue i as post p_k is beingshared.
$\Delta S(a_x, a_y, i, p_k)$	shift in stance (of a_x) due to a sanction (by a_y) for p_k it shared on the issue i.
POV(a_x, p_k)	POV of a_x after p_k has diffused in the social network
neighbor(G, a_x, p_k)	all neighbors of a_x in the social network G which receive p_k from a_x
numAgents(G)	Total number of agents in the social network G

Table E.2: Notations used to describe the simulation design.