# MOVING PAST THE IDEOLOGY:

## How Do Violence and Popularity of Social Movements Affect the US Liberal News Coverage?

by

Dauren Koptleuov

Submitted to

Department of Political Science
CENTRAL EUROPEAN UNIVERSITY

*In partial requirements for the degree of Master of Arts*

Supervisor: Mariyana Angelova

Central European University, Political Science Department

Vienna, Austria
2020

# Abstract

News media coverage is commonly regarded as one of the key means for success for any social movement (Gamson and Wolfsfeld, 2013). However, the relationship between these two political actors is rather complicated. Scholars observed a trend in which the US news media outlets tend to delegitimize a movement through framing, sourcing, and other journalistic practices (Gitlin, 1980; McLeod and Hertog, 1999, Kilgo and Harlow, 2019). This tendency is referred to as the protest paradigm in communication studies. More recently, the growing polarization within the US media circle challenged the paradigm: the matching ideology of media and social movements was observed to reduce the negative bias in the coverage (Weaver and Scacco, 2013; Kim and Shahin, 2020). Up to date, the scholars of media and social movements seemed to be focusing on how news cover protests, but not why. This study analyzes the New York Times (NYT) coverage of the Black Lives Matter (BLM) in the first two weeks of the George Floyd protests. According to the ACLED US Crisis Monitor, the first week of the protests was reported to have much more incidents of violence than the second (17.23 percent versus 1.64). This coincided with the growing popularity of the Black Lives Matter movement in the same time frame. This thesis addresses how violence and popularity affect the tone of the news coverage, namely its negativity, and aims to contribute to the theories that challenge the protest paradigm phenomena. The negativity of the coverage is analyzed through framing and sourcing. The first refers to the number of delegitimizing frames, while the second refers to the number of quotations of authority-related public officials on the topic of protest. Due to the skewed distribution of response variable values, this thesis employs logit models and OLS regression models. Logit models aim to predict the odds of the NYT to be negative or not in their coverage. OLS regression models seek to test the possible linear relationship between the violence and popularity and the degree of negativity of NYT coverage of George Floyd protests. The findings reveal that the time passed since the first day of protests has a more profound effect on the odds of NYT being negative in their coverage than expected. Violence shows a modest positive effect on the degree of negativity through framing, while the support for Black Lives Matter slightly diminishes the chances of NYT to quote an official.

## Acknowledgements

I would like to express my sincere gratitude to the Political Science Department for the support, cooperation, and understanding in times of crisis. I am thankful to all my friends (old and new ones, whom I was lucky to meet at CEU), my small, but loving family and of course, Catherine. Thank you for always being there for me.

Finally, this thesis would not be possible without my supervisor, prof. Mariyana Angelova. Thanks for always finding the time to reassure and guide me through the scary stuff (statistics).

# Table of Contents

# List of Figures

# List of Tables

# List of Abbreviations

**BLM –** Black Lives Matter

**NYT –** New York Times

**ACLED –** Armed Conflict Location and Event Data

**OLS –** Ordinary Least Squares

# Summary Statistics Disambiguation

**Support –** the percentage of US citizens expressing their support/approval for the Black Lives Matter based on their answer the question "Do you support or oppose the Black Lives Matter movement?"

**Neutral –** the percentage of US citizens holding neutral attitudes towards the Black Lives Matter based on their answer to the question "Do you support or oppose the Black Lives Matter movement?"

**№ of Events –** the number of protest events in a single day in a particular state

**Learning –** the time passed in days since the first day of George Floyd protests as part of nation-wide Black Lives Matter protests in 2020

**Violence –** the percentage of violent events extracted from the total number of protest events in a single day in a particular state

# Chapter 1: Introduction

This chapter of the thesis aims to familiarize the reader with the context of the George Floyd protests in 2020. Apart from that, it provides insights on the relationship of media with social movements, the role of ideology, and how other factors might affect the tone of the coverage social movements might receive from media. Finally, I present a research question of this thesis, a brief roadmap of the argument, and its normative implications.

## 1.1 Context

The year 2020 had been keeping American newsmakers busy. First, the tensions erupted between the US and Iran (Atwood, 2021), then, the US Senate attempted to impeach then-president Donald Trump (Politico, 2020). Finally, these events were followed by the deadly outbreak of COVID-19 (Financial Times, 2020). All of this took place while the citizens were getting ready for the upcoming presidential elections. However, as summer was approaching, these and other major political news events got overshadowed by the consequences of George Floyd's murder in Minneapolis, Minnesota. Floyd was an African-American who died in the custody of four police officers. In the video made by one of the witnesses, Floyd was seen pinned to the ground by a white male officer's knee on his neck, crying and struggling to breathe, and eventually passing out of suffocation (BBC, 2020). Shortly after such a shocking display of violence became viral on social networks, the streets of Minneapolis were flooded with demonstrators (NPR, 2020). The outrage with the growing number of cases of systemic racism, police brutality, and social injustice against the non-white community has resulted in protests becoming nationwide in just a week (LA Times, 2020).

Even though the George Floyd protests as part of the wider Black Lives Matter movement were already gaining momentum on social networks, it was also important for the protesters to spread the word through any other alternative sources, including the news media (The Guardian, 2020). The officers involved in the murder were not immediately prosecuted, so spreading public awareness about the George Floyd's case was one of the key aims of the demonstrations (The Guardian, 2020). As the primary platform which can amplify the voices of protesters so that they reach the public (Gamson and Wolfsfeld, 1993), news media attention is commonly regarded as a key to the success of social movements. Despite the growth of social media as a source of information in recent years, the news media is still widely seen as one of the main platforms of attracting public attention to the protests (Kilgo and Harlow, 2019). In the past, the newsmakers showed a tendency to ignore and/or misrepresent progressive social movements and protesters (Gitlin, 1980). The systematized observations of such tendencies evolved in a theory known as the protest paradigm in communication studies (McLeod and Hertog, 1999).

The theory proposes that in response to social movements challenging a societal status quo, the news media adopt biased journalistic practices that make social movements appear in a rather negative light before the public (McLeod, 2007). However, with the rise of the Internet as the new primary medium of information, the traditional media had to adapt to maintain their presence online. Alongside these transformations, the news media domain was observed becoming more ideologically polarized than ever (Garrett et. al, 2016). This meant that the Left-leaning social movements gained new hope to be adequately represented in the mainstream news, the online domain of which was becoming more and more diverse (Pew Research Group, 2014).

These developments are consistent with the recent theoretical challenges to the protest paradigm. The arguments in favor of reconsidering the protest paradigm pointed at the importance of matching ideologies of a media outlet and a social movement. Such match between the ideologies seemed to be a deciding factor when it came to predicting the tone of the coverage (Weaver and Scacco, 2013; Kim and Shahin, 2020). Weaver and Scacco (2013), in their study of the Tea Party Movement, observed that the conservative movement managed to receive a mostly negative, delegitimizing coverage from the liberal media outlets, while Fox News produced much less negativity in their narratives. Empirical academic (Kilgo and Harlow, 2019) and journalistic (The Guardian, 2020) observations confirmed that the theory works in the opposite direction as well: in summer 2020, the conservative media tended to cover the Black Lives Matter movement in a highly negative way.

Because of ideological differences, attitudes towards progressive social movements, and a history of strained relationships in the past (Kilgo and Harlow, 2019; Kilgo and Mourao, 2019) it seemed that it would be unfeasible for BLM to rely on the coverage of the conservative media to provide the sufficiently fair coverage of the struggle against racism and police brutality. A common example would be Fox News that in line with then-President Donald Trump adopted the narrative that viewed the Black Lives Matter movement and its followers as "looters", "rioters", and "domestic terrorists" (The Guardian, 2020). Therefore, in the case of the Black Lives Matter movement, the protesters would have to rely on the support of liberal (or left-leaning, in the US political context) news outlets in hopes that it would cover the protests in a relevant manner.

## 1.2 Timing and Dynamics of the Protest

The first two weeks of the George Floyd protests marked the beginning of the nationwide Black Lives Matter movement in 2020. Those two weeks provided several prominent trend in the protest dynamics. The first week of protests, May 26th to June 1st, was arguably the most "violent", according to the Armed Conflict Location and Event Data (ACLED) project (2020). 17.23 percent out of total 1928 protest events resulted in the episodic manifestations of violence, either against individuals, public or private property. The second week (June 2nd to 8th), however, was by far more peaceful: only 1.64 % resulted in violence out of 3386 protests in total (US Crisis Monitor, 2020). This surprisingly coincided with the growing public support of the Black Lives Matter movement after the first week. It started with 45% of the US population strongly supporting BLM right after George Floyd was murdered, which peaked on June 1st with 52% and has remained such in week two (CIVIQS approval metrics, 2020).

Bar plots in figures 1.1 and 1.2 demonstrate the protest dynamics of the initial research sample used for the analysis (n=192). The y-axis shows mean percentages of violent events (figure 1.1) and US citizens expressing their support (figure 1.2) over the time period of the first two weeks of George Floyd protests on the x-axis. As the graphs reveal, the observations in the sample followed the similar dynamics. The mean percentage of violent events across the US experienced a drastic decrease after the first week: it dropped from around 50% to less than 10%. At the same time, the mean percentage of US citizens supporting Black Lives Matter increased slowly but steadily. For day 13, however, the values for the support were missing, and on day 14 the percentage experienced a slight decrease. This was primarily due to the low number of observations in the last two days, as can be seen in figure 1.3, which shows the number of text reports of the New York Times on the George Floyd protests. This resulted in the several data transformations, which are discussed further in the research design chapter.

Figure 1.1 Protest dynamics for violence during protests
expressed in the mean percentage of violent events across states (y-axis). The x-axis represents the time period of the first two weeks of the protests, May 26 - June 8.

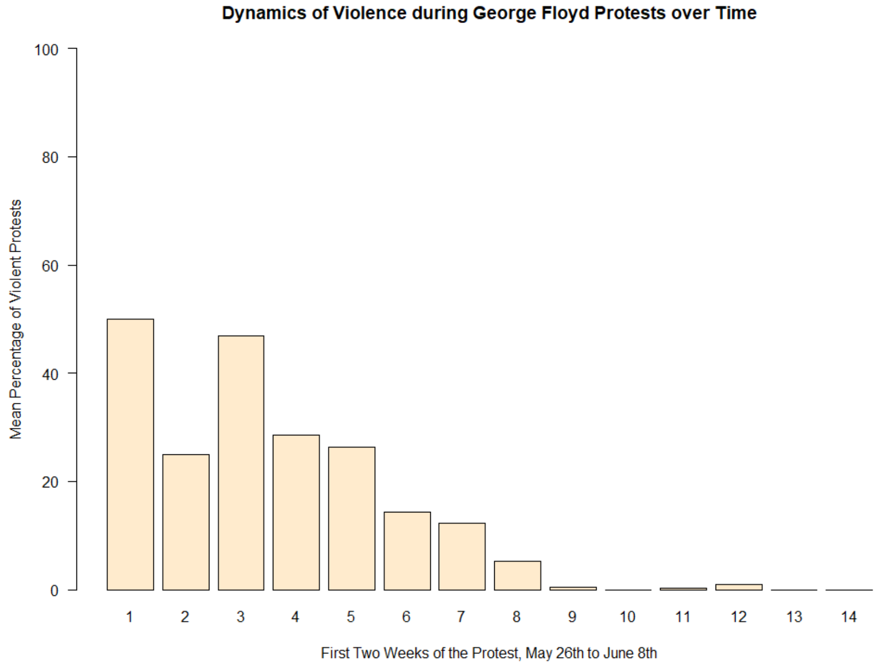



Figure 1.2. Protest dynamics for popularity during protests
expressed in the mean percentage of violent events across states included in the sample (y-axis). The x-axis represents the time period of the first two weeks of the protests, May 26 - June 8.

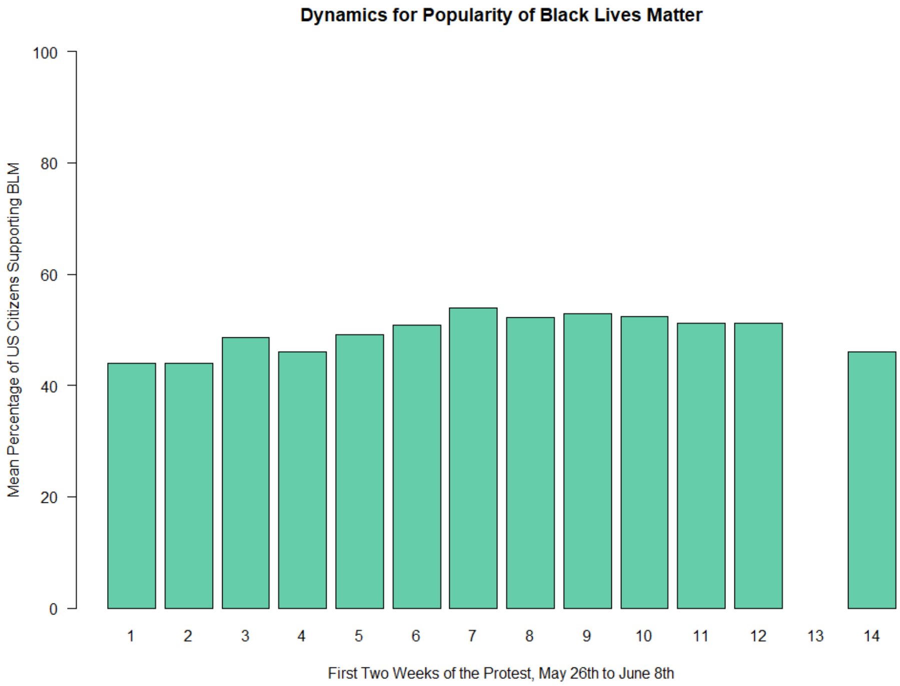
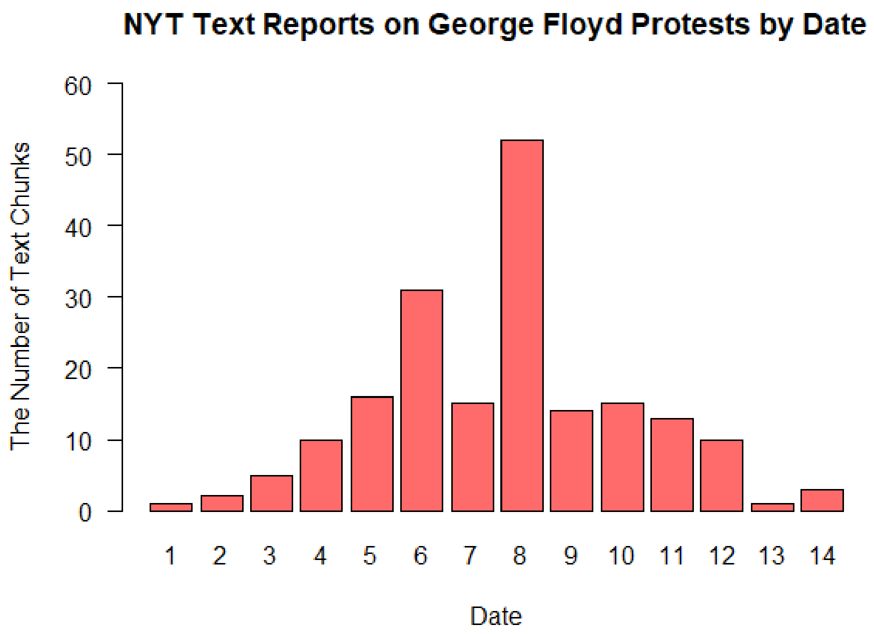


Figure 1.3. The number of NYT reports on protest activity over the first two weeks of George Floyd protests, May 26 - June 8.




Such fluctuations in the protest dynamics as growing popularity, in spite of the amount of violence during the first week, were not solely due to the impact of the news media. The Black Lives Matter movement has been observed to have a strong presence on social media platforms, like Twitter, Facebook, Instagram, and others (Mundt et. al, 2018). Additionally, the relationship between news media and public opinion is not as straightforward as it seems. According to Strömbäck (2012), the media are able to not only influence, but also to reflect public opinion, especially after the opinion polling became widely available to the newsmakers via Internet. Multiple studies on the media coverage of the economy had argued the possibility that the media coverage follows public opinion on the issue to maintain its audience, which indicates that the shifts in social trends and attitudes affect the media coverage more than it has been intuitively thought it does (Hopkins et. al 2017). In this thesis, I expect this tendency to be also observed in the liberal media coverage of protests and social movements.

## 1.3 Research Question

As mentioned previously, scholars established that the conservative media outlets tend to cover Black Lives Matter negatively by adopting delegitimizing frames as well as sourcing practices that misrepresent the movement (Kilgo and Harlow, 2019). This was confirmed by the recent theoretical challenges to the protest paradigm that point out that mismatching ideologies of an outlet and protests would most likely lead to increasingly negative media coverage. Nevertheless, in this thesis I am willing to extend that theoretic claim. I theorize that there were empirical factors that drove the tone of the media coverage, even if the ideologies of an outlet and a social movement matched. Due to the puzzling interplay of escalating violence and increasing popularity, it was interesting to probe into the reasons behind the tone of the liberal US news media coverage of George Floyd protests as part of the wider Black Lives Matter (BLM) movement in 2020. I investigate the recent coverage of George Floyd protests around the US by the New York Times (NYT). The New York Times is known as one of the most popular (Elmasry and El-Nawawy, 2017) and reputable left-leaning (in the context of American politics) media outlet with "consistently liberal" and "somewhat liberal" as the larger part of their audience (Pew Research Group, 2014). Given the puzzling observations in the protest dynamics, I move forward with the following research question:

**How do violence and the popularity of social movements affect the news coverage by liberal media outlets?**

In this thesis, I propose that the liberal news media in their coverage of the progressive social movements (like BLM) adapt to the new dynamics of the protests, namely the decreasing violence and increasing popularity. I am committed to investigate how both violence (measured by the percentage of violent protest events) and popularity (measured by the percentage level of civic approval online) could influence the negativity of the New York Times coverage of the first two weeks of the BLM protests. I would expect that the negativity would increase if violence during protests increases. If the popularity of the BLM increased, however, I would expect the negativity to drop. In this research, the negativity is measured by two proxies:

- framing, or the number of negative frames used in relation to protesters and their activity,
- and sourcing, or the number of quoted authority officials on the topic of protest.

I expect that if the coverage would shift the degree of negativity in accordance with changing dynamics of the protests (less violence, more popularity), then it would mean that the protest paradigm would not persist. Alternatively, if the negativity kept increasing despite decreasing violence and growing popularity, it would mean that the NYT conformed to the protest paradigm. This would be an especially puzzling result, since it would contradict the previous challenges to the protest paradigm based on the ideological alignment (Weaver and Scacco, 2013; Kim and Shahin, 2020), as well as theoretical expectations developed in this thesis. I expect that the tone of the news media coverage of the nation-wide protests is still sensitive to the scale of violence and popularity of the social movement, even if its ideology matches with an outlet providing coverage.

## 1.4 Normative Implications

The frequent usage of negative frames, as well as quoting officials predominantly instead of protesters, forms an increasingly negative media narrative of social movements (McLeod and Hertog, 1999). This leads to further delegitimization of mostly peaceful protesters by labelling them as criminals and rioters and contributes to easier justification of the use of force against them (Kilgo and Harlow, 2019). A more recent study by Kilgo and Mourao (2021) showed that exposure to negatively biased media narratives contributed to more negative perceptions of the Black Lives Matter movement and protesters.

Such anti-protest bias in the media coverage is problematic for the liberal democracy. Instead of assisting social movements to address public grievances and keep the government in check by "promoting necessary social changes" (Meyer, 2003), it rather marginalizes and delegitimizes them, and at the same time legitimizes the repressive measures against the protesters (Gitlin, 1980). It would be especially problematic if the liberal media outlets engaged in such practices, often being the only mainstream option for progressive protesters to spread their message and share their grievances. While social protest is one of the essential parts of the liberal democratic ecosystem, its systemic delegitimization and marginalization in the mainstream media presents a fundamental challenge to liberal democratic values as well as the journalistic professionalism in the US.

# Chapter 2: Literature Review and Theoretical Framework

In this chapter, I go over the existing theories and recent challenges to the protest paradigm. Further, based on the literature, I develop a theory of how empirical factors, such as violence and popularity might affect the negativity of the media coverage of George Floyd protests in 2020. Lastly, I come up with four hypotheses that state the expected relationship between predictors and the response variable, measured through framing and sourcing.

## 2.1 News Media, Social Movements and Protest Paradigm

The literature on the relationship between news media and social movements holds a dominant view in which the two are closely interrelated (Molotch, 1979; Gitlin, 1980). According to Gamson and Wolfsfeld (1993), both news media and social movements share a mutual interest in each other, since social movements make content for the news, and news media can spread the desirable message and contribute to mobilization for social movement groups. The advocates for social change argue that progressive social movements take on the institutional challenge of checking on and holding the governments accountable. Despite often being short-lived, progressive social movements press for addressing the necessary social changes, be it universal suffrage, limiting military interventions, combating austerity, or other civil rights issues (Meyer, 2003). Meyer (2003, p. 32) argued that to influence public policy, social movements must bring the attention of the public to the goals and purposes of the struggle. In the context of media and movements, news media is therefore considered as the key communication mediator between the protesters and the public as well as the governing actors (Tenenboim-Weinblatt, 2014). To put it in more exact terms, "protest leads to media coverage of protest events, which leads to media coverage of issues relating to the protest more widely, which leads to politics" (Vliegenthart et. al, 2012)

In the past, however, the newsmakers had shown a tendency to delegitimize progressive social movements and protesters (Gitlin, 1980; McLeod and Hertog, 1999), making it appear less favorable in the eyes of the public. More recent literature findings demonstrated that the media tend to resort to the use of delegitimizing frames and other methods, which contributes to misrepresenting the essence of protesters' identity, their goals, and demands, and the core of a social movement they advocate for (Boykoff, 2006; McLeod, 2007; Boyle et. al, 2012). These observations collectively contributed to a theory referred to as **the protest paradigm** in communication studies. The argument at the core of the protest paradigm suggests that the US news media covers domestic protests in a predominantly negative manner through a variety of journalistic practices (McLeod and Detenber, 1999; McLeod, 2007; Weaver and Scacco, 2013). The most prominent of them include:

- active use of frames that stress the negative side of the protest activity (e.g., describing the whole protest events as "riots", "mayhem" and other delegitimizing language expressions).
- over-reliance on sourcing authority officials that are, at best, unfamiliar with the movements' goals and demands and openly hostile towards protesters at worst (e.g., quoting mayors, etc. and other elected officials' comments on the situation).

Gitlin (1980), in his work on the progressive social movements and the media coverage, pointed out that in the US, the corporate media domain is sometimes inseparable from the governmental domain, which strains the attempts of protesters to collaborate with the media. This suggests that while protesting the capitalist establishment, the progressive groups attack fundamental values of both government and media structures, which results in coverage that is driven by a negative bias. Gitlin (1980) elaborated that such established order dominates the public narrative by producing media frames that are extremely challenging for protesters to contest.

More recent works in the studies of media coverage of protests and social movements referred to such order as the "status quo" (Kilgo and Harlow, 2019). Kilgo and Harlow (2019, p. 510) outlined that for progressive social movements, the established power relations as "white supremacy, patriarchy, capitalism, and ableism" can be collectively understood as status quo in relation to the confronting stance of the protesters against social injustice. Feminist protests, therefore, are aimed at challenging the patriarchal state of society. Anti-racist movements are focused on challenging white supremacy. The overall purposes of a progressive social movement might include contesting multiple elements of the existing status quo. For example, George Floyd's protests combine the struggle against systemic racism and police brutality, while the wider Black Lives Matter movement also amplifies the voices of Black women and transgender people (Tillery, 2019).

## 2.2 Ideological Alignment and Tone of the Media Coverage

Recent works in media and social movements studies presented an intriguing theoretical challenge to the protest paradigm. It followed a rather intuitive critique that became more relevant when US media had grown more and more polarized with the rise of the Internet as a key medium of information (Garrett et. al, 2016). The findings suggested that the US news media covered the movements in line with their ideological stance in relation to both domestic (Weaver and Scacco, 2013) and international protests (Kim and Shahin, 2020). One interesting example was the study of mainstream news coverage of the Tea Party Movement, which authors defined as a right-leaning, conservative movement (Weaver and Scacco, 2013). The evidence showed that the liberal media resorted to profoundly more negative coverage of the movements with opposing ideologies, i.e., Tea Party Movement in the context of Weaver and Scacco's study. The degree of negativity in the coverage was assessed by the presence of "marginalization devices" that were defined as frames used to construct negative, delegitimizing images of the movement (Weaver and Scacco, 2013).

Fox News, on the contrary, was reported to use significantly fewer "marginalization devices" on the average per script (0.83) as compared to the liberal media outlets, like CNN (1.48) and MSNBC (2.16) (Weaver and Scacco, 2013, p. 74). In the case of Black Lives Matter, previous theoretical findings suggest a similar pattern on the part of ideologically opposing outlets. Kilgo and Harlow (2019) observed the right-leaning Texan media cover the Black Lives Matter protests in an increasingly negative way. For conservative media, like Fox News, the anti-police agenda of the Black Lives Matter movement would be a direct attack on one of its values. In the past, Fox publicly supported the Blue Lives Matter group, a counter-movement that rejects BLM agenda and fights for the harsher prosecution of those involved in the killings of police officers (Washington Post, 2020). And in the present, the journalistic observations on the Left have demonstrated that the overall tone of the Fox News was consistently negative towards the Black Lives Matter movement and its supporters by referring to the protesters as "looters", "rioters", and "domestic terrorists" (The Guardian, 2020).

There were, however, some essential unmentioned limitations in Weaver and Scacco's theory. The authors did not probe into confounding factors that could motivate the media coverage apart from the ideological matching of an outlet and a social movement. They suggest that the studies of media coverage should not refer to the protest paradigm as such if there is evidence of negativity in the coverage "falling along the lines of party or ideology" (Weaver and Scacco, 2013, p. 78). This did not respond to the complications established by theoretical revisions by the scholars who introduced the protest paradigm as a concept. One notable revision pointed at how tactics of the protests matter more for the tone of the coverage than their goals (Boyle and Armstrong, 2009).

The previous findings in the protest paradigm theory (McLeod and Hertog, 1999; McLeod, 2007) pressed to attach a greater significance to both the goals and tactics of the protests. However, Boyle and Armstrong (2009) argued that the goals and tactics should be considered separately while analyzing the media coverage of the movements. The goals of the protests were defined in terms of the relationship of the protests to the status quo, i.e., whether a movement aims to preserve it or oppose it aiming for an immediate social change (Boyle et.al, 2012). The tactics of the protests were referred to as the methods through which protesters aim to achieve their goals (Boyle et.al, 2012). Subsequent arguments for the analytical separation of the two terms contributed to the evidence that the tactics of the protests play a more significant role than the goals (Boyle and Armstrong, 2009). In their study of abortion protests, Boyle and Armstrong (2009) found that the protesters' goals did not affect the tone of the coverage as significantly as their tactics did. In other words, the ways (i.e., tactics) in which protesters chose to express their grievances and the radicality of their actions proved to affect the tone of the coverage more prominently than protesters' position on abortions (either pro-life or pro-choice). Another study by Boyle et. al situated in the international media context found a similar pattern. The authors discovered that those protest groups who employ radical tactics are treated with significantly more negative coverage irrespective of their goals and the outlet's ideology and location (Boyle et. al, 2012). These findings set forth a contradiction unaddressed by Weaver and Scacco's theoretical challenge to the protest paradigm. This study aims to address this gap by testing how empirical factors, such as violence and popularity might affect the negativity of the New York Times coverage of George Floyd Protests.

## 2.3 News Media and Black Lives Matter



In the past, the issue of race in the US news media seemed to be falling under institutionalized stereotypes, as the news tended to disproportionately portray Black Americans as criminals and White Americans as crime victims, as opposed to what descriptive statistics suggest (Dixon and Linz, 2000a; 2000b; Dixon and Acozar, 2006). The previous inquiries in the coverage of Black Lives Matter protests in 2016 provided similar findings: the tone of the news was mainly consistent with the ideology of a news outlet (Elmasry and El-Nawawy, 2017).

Conservative news outlets were observed to align with the protest paradigm in a rather specific way: the study by Kilgo and Harlow (2019) identified that Texan right-leaning newspapers follow a social hierarchy pattern, in which some of the protests for social justice are being portrayed more negatively, than the others. The study indicated that BLM protests were portrayed in an increasingly negative and delegitimizing way, as compared to feminist protests, environmentalist rallies, protests on immigration rights, and the conservative movements (Kilgo and Harlow, 2019). One important finding showed that negative media framing contributes to growing criticism of the movement and protesters as well as reinforcing the preexisting negative stereotypes about people of color (Kilgo and Mourao, 2021).

Another study on the coverage of Black Lives Matter (Elmasry and El-Nawawy, 2017) was aimed at looking into the coverage of one local left-leaning newspaper (St. Louis Post Dispatch) and one nation-wide outlet (the New York Times). The authors expected that the media would negatively cover the protesters due to the several violent clashes with the police at the beginning of the protests. This expectation seemed to echo Boyle and Armstrong's (2009) suggestion to pay closer attention to the protesters' tactics, as the violence typically escalates the conflict between protesters and authorities and therefore increases the chances of gaining media attention (Gottlieb, 2015). Yet, according to Elmasry and El-Nawawy (2017), the analyzed news outlets were reported to provide a rather "sympathetic" coverage. The findings demonstrated the moderate usage of negative frames as well as adopting such sourcing strategies that quote protesters much more frequently than the officials (Elmasry and El-Nawawy, 2017).

## 2.4 Theory Building

It seems that even though the issue of institutionalized racism and the Black Lives Matter movement became much more recognized than before, a considerably large part of the public still associated protesters with illegal activity and violence (Pew Research Center, 2016). The

Armed Conflict Location and Event Data (ACLED) project's report on BLM protests in May-August 2020 revealed that out of more than 7000 protest events, 93 % were peaceful. The other 7% composed of such illegal activities as looting, rioting, clashes, and other violent acts. Despite what these statistics show, a poll from private data investigation company Morning Consult showed that 42% of Americans think that the Black Lives Matter protesters tend to "incite violence or destroy property". Such mismatch between the empirical statistics and public perceptions of the movement is largely explained by the "biased media framing", the ACLED report stated (2020, p. 6).

Both journalistic (The Guardian, 2020) and academic (Kilgo and Harlow, 2019) investigations demonstrated that the conservative media outlets tend to cover Black Lives Matter in a predominantly negative and delegitimizing way. In this thesis, I aim to check whether liberal media follow a similar media coverage tendency on the example of the New York Times. More importantly, however, I am interested in seeing whether the tone of the coverage by NYT can be motivated by empirical factors, like violence during the protests and the popularity of the movement. The choice for the following builds on the intuition of Boyle et. al (2012), which stresses the importance of protest tactics over goals. I theorize that both violence and popularity are directly related to the tactics of the protest. Violence demonstrates whether the movement resorts to radical measures, while popularity points at the attempts to appeal to the wider American population through a variety of peaceful means.

As previous works on the protest paradigm (Kilgo and Mourao, 2019; 2021) established that media coverage can influence the reader's perceptions of social movements and protesters, this thesis does not investigate the media effects as such. Instead, as Weaver and Scacco (2013, p. 63) put this, I focus on "the content of the frames in and of themselves". Frames, in the context of this study, conceptually refer to framing as one of the three classical models of political communication (Scheufele and Tewksbury, 2006). As opposed to priming and agenda-setting, framing's communicative power is in the meaning attached to the media characterizations of the issue. These meanings can carry positive and negative meanings, and as psychosocial research demonstrated, the media choice on the tone of these meanings can influence the readers' understanding as well as attitudes on an issue (Scheufele and Tewksbury, 2006).

Protest paradigm scholars view framing as the essential journalistic practice that drives the tone of the coverage, indicating the presence of the protest paradigm (McLeod, 2007). In this regard, framing is often investigated along with sourcing. Sourcing is a journalistic practice indicating media outlet's preferences of quoting individuals belonging to various groups (mostly protesters and officials) in their news coverage (Kilgo and Harlow, 2019). It can directly affect the tone of the coverage through the following preferences:

- giving voices to protesters can lead to the more sympathetic coverage, as it provides the closer, legitimizing insight into the social struggle (Bray, 2012).
- quoting official sources (police, government, local politicians) tends to contribute to delegitimizing narrative through the lack of familiarity of a movement's purpose, but more so because of antagonizing nature of the relationship between those in power and those who contest the power through protests (McLeod, 2007).

In this thesis, I address the puzzling question of whether violence and popularity could influence the tone of the media coverage of a social movement with matching ideology. The literature demonstrates that most of the previous works on the protest paradigm were carrying the exploratory research purpose, i.e., how media covers progressive social movements. While those findings provide interesting descriptive implications, in this study I aim to pursue the confirmatory approach that could provide the answers to why media cover movements in the ways they do. Case-wise this work is similar to Elmasry and El-Nawawy (2017), as it also probed into the NYT's coverage of Black Lives Matter, yet this study proposes several elaborations both methodologically and theoretically.

First, unlike Elmasry and El-Nawawy, I organized the units of analysis (text coverage) on a state-date level, which allows me to track the changes in the negativity of coverage on specific spatial and temporal levels. Second, organizing data in such a manner allows tracing whether the tone of the coverage correlates with such empirical factors as violence and popularity of the movement, which was missing from previous works on media coverage of the protests. This is a crucial distinction, which would benefit the more concrete understanding of the reasons behind the tone of the news coverage, namely its negativity. Comparing the negativity to the actual numbers from empirical statistics could point towards the extent to which the media outlets might be biased, regardless of whether their ideology matches with the ideology of their object of coverage. Finally, in this study, I am not interested in the numbers of negative frames or quoted officials per se, but rather in seeing whether the variation in those numbers can be explained by the linear relationship with violence and popularity.

Therefore, apart from providing a more elaborate methodological approach, this study offers a new theoretical perspective as well. Complimenting the recent theoretical challenges to the protest paradigm, I concur that ideological alignment plays a significant role in the tone of the coverage. However, taking into consideration theoretical revisions by Boyle et.al (2009, 2012), the ideology factor might become of secondary importance to the tone of the coverage in the presence of empirical factors related to the protest tactics: violence and popularity. In this paper, I argue that such empirical factors related to the protest tactics as violence and popularity can play a more significant role:

- if the protests are getting more violent, then the tone of the coverage increase the negativity of its, regardless of matching ideologies.

- if the protests are receiving growing wide support from the population, then the tone of the coverage decreases negativity , regardless of matching ideologies.

## 2.5 Hypotheses

As mentioned previously, in this thesis I am committed to test whether or not the violence and popularity of Black Lives Matter affected the tone of the coverage of the New York Times. In my first hypothesis, violence during the protests is an independent variable (or predictor) measured with the percentage of violent events taking place in a particular state on a particular date. The negativity of the coverage is a dependent variable (or response), which is measured through number of negative frames, or framing and the number of quoted officials, o sourcing.

- H1a: An increase in violence causes an increase in the number of negative frames.
- H1b: An increase in violence causes an increase in the number of quoted officials.

The data on violence is provided by the US Crisis Monitor (ACLED, 2020), which can help to pinpoint the total percentage of any manifestations of violence during Black Lives Matter protests. It can be narrowed down to specific periods or even protest events. From the text, via automated text analysis, I extract two proxies that measure the tone of the coverage, my dependent variable: delegitimizing frames (e.g., rioters as opposed to protesters) and quoting the officials and police more frequently, than the protesters.

The same procedure follows in the case of my second independent variable, which is the popularity of the social movement (in this case, Black Lives Matter). The popularity variable is measured by two proxies: the percentage of people expressing support for BLM (or support), and the percentage of people having neutral attitude towards BLM. The percentages were extracted from CIVIQS approval metrics. The poll reflects a public perception of Black Lives

Matter, which can provide a representative sample of respondents that support, condemn or are neutral to the movement measured on daily basis.

- H2a: An increase in popularity of the movement causes a decrease the number of negative frames in the coverage.
- H2b: An increase in popularity of the movement causes a decrease the number of quoting of officials in the coverage.

# Chapter 3: Research Design

This chapter discusses thesis methodology and other research design decisions regarding case selection, data collection, measurements, and operationalization techniques. In the first section of this chapter, I explain the rationale behind choosing two first weeks of the protest as the time frame of this research. Additionally, the first section justifies the choice of the New York Times as a typical case of the US liberal media outlet with ideology and popularity of the NYT as two main reasons.

## 3.1 Time Frame and Case Selection

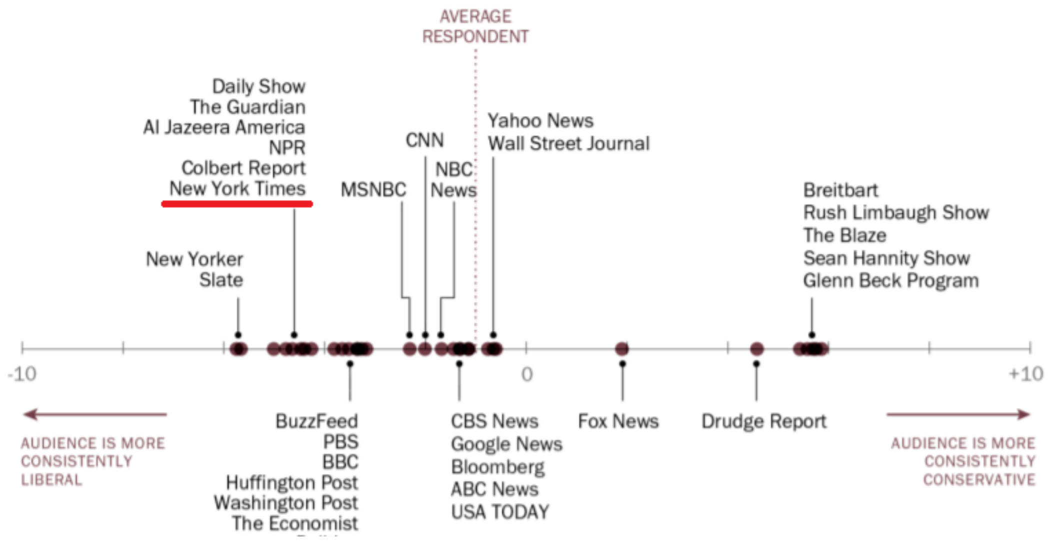
As stated in the Introduction chapter, the time frame for this thesis project was pinpointed to the first two weeks of George Floyd protests in 2020. Since the liberal media were previously observed to change their coverage of progressive protests in the long term by changing their framing focus (Gottlieb, 2015), one intuitive strategy was to pick a period that is temporally close to the start of protests. That way, it would be possible to observe the coverage unaffected by the shifts in public opinion. During the preliminary research stage, two interesting observations were discovered. In the first week of protests, the violence erupted on the American streets: according to ACLED US Crisis Monitor (2020), 17.23% out of 1928 protest events were violent in the first week of protests of Black Lives Matter protests. However, the situation settled down fast before the start of the second week: only 1.64% of 3386 protests resulted in violence (ACLED, 2020). Despite the damage caused by the intense first week, the nationwide popularity of the Black Lives Matter movement was on the rise. It was 46% of the US population expressing their support for BLM until May 30th, when it started growing, peaking on the second week of protests, June 2nd with 53% (CIVIQS, 2020). It remained on that level until June 8th, marking the end of the second week and being the all-time high for BLM protests in 2020.

These changes in protest dynamics were interesting phenomena to observe by themselves: it seemed that the popularity seemed to increase despite the violence in the first week. Such gains in popularity, in turn, could potentially explain the consequent decrease in violence over the second week (one could speculate, for example, that having more citizens on the streets could reduce radical sentiments among protesters). Nevertheless, these observations served only an instrumental purpose for this thesis. Due to the different research goals, I did not investigate the reasons behind the changes in protest dynamics but rather used them as a reference point. I am primarily interested in seeing how changes in violence and popularity would affect the tone of the New York Times coverage of George Floyd protests. I believe that narrowing the time frame of the research to those two weeks was desirable not only because of rapid changes in violence and popularity but also because the effect of that variation could flatten out as more time goes by.

Regarding the media outlet selection, the New York Times seemed to be not only one of the most popular US media outlets among the population but among the academics as well. Prior to this study, NYT had been scrutinized by numerous authors studying news media coverage and its effect on progressive social movements like Black Lives Matter (Wright and Reid, 2011; Gottlieb, 2015; Elmasry and El-Nawawy, 2017 and others). As I mentioned earlier, I am committed to extending the challenge to the protest paradigm already established by the ideological alignment argument (Weaver and Scacco, 2013; Kim and Shahin, 2020). Thus, in my research, I am specifically interested in liberal (or left-leaning, in the US context) media news outlets. In the case selection procedure, I followed Gerring's (2008) guidance on the typical case, which would be representing the whole group (US liberal media outlets, that is). For the criteria of typicality, I selected two main features: ideology and popularity of an outlet.

For ideological standing, I referred to the Pew Research Group's report on the ideological scaling of the US media (figure 3.1). According to Pew (2014), the New York Times is one of the top five liberal outlets when it comes to popularity and trust among consistently liberal readers. It also shared one of the largest sources of distrust among the conservative readers (Pew Research Group, 2014). These ideological preferences demonstrated that the New York Times can be safely considered as a typical liberal news outlet. The following media ideological scaling graph in figure 3.1 confirms it, locating NYT (underlined in red) straight in the left-leaning middle:

Figure 3.1. Ideological Placement of US Media Outlets
(Pew Research Group, 2014).




Apart from the ideological placement, another important criterion was the readership numbers of an outlet – the more popular the outlet is, the more impact it can have on its audience. Watson (2021) reported that around 2016 the New York Times had a second highest newspaper circulation in the world. As most of the printed newspapers faced the need to turn digital, NYT reported having five million paid subscribers in the fourth quarter of 2020 in their business transparency press release (The New York Times Company, 2021). It had been placed in the top-3 brand in "online brands" and top-10 in "TV, radio and print" categories by the Reuters Institute in their massive report on the state of Western media (Reuters Institute, 2020).

Thus, satisfying both criteria, the New York Times presented a typical case for a liberal news media outlet in the US.

### 3.2 Variables

In this section, I go over the list of variables selected for my model. In short, there are two independent variables in this study: violence during the protest events and popularity of the social movement. I theorize that both could influence the tone of the coverage, regardless of whether the ideologies of an outlet and a social movement matched or not. Both violence and popularity were noted after becoming familiar with background protest dynamics in the preliminary research stage. The measures for both violence and popularity were based on the data from empirical statistics.

In the case of violence, the variable was measured as a percentage of the BLM-related violent protest events to the total number of BLM-related protest events in the first two weeks of George Floyd protests in 2020. The information about these events was provided by the extensive ACLED US Crisis Monitor database on political violence and conflicts in the US. The database was manually compiled by a team of experienced researchers, who update it on a weekly basis. They extracted the information on political violence and protest activity from secondary sources, such as local and nationwide news reports, and after a close inspection, they coded each protest event as an "atomic unit". Such coding helped other researchers to determine the actors, their motives, and the exact date and location of the event.

The measurement for popularity variable was designed to reflect the levels of support for the movement among the American population. This data was provided by the CIVIQS polling group that is the only widely available database that can demonstrate US citizens' perceptions on the variety of political topics. It gathered polling data from online users on daily basis. They cited increased Internet availability as one of the key reasons for choosing to survey US citizens on the Web – according to Pew Research Group, almost 90% of Americans have access to the Internet.

While it was possible to "match" the data from both CIVIQS and ACLED on a temporal level (day-by-day), it still required some reorganization, when it came to spatial considerations (state or city?). While each of the protest events had its own location (which made it possible to code them as "atomic units"), the perceptions of citizens on the Black Lives Matter were revealing the support on the state level only. Thus, the independent variable popularity was only available for collection on the state-date level. As the research purposes of this thesis dictated the variables and unit of analysis to be matching both temporally and spatially in the unified dataset, it was decided to organize violence on the state-date level as well. During the analysis, the popularity variable was composed of two proxies, support and neutral. The former indicated the state-wide percentage of US citizens supporting the Black Lives Matter, and the latter represented the percentage of those US citizens with a neutral attitude towards the BLM.

In this thesis, the negativity of the coverage was selected as the response variable. It was measured by two proxies commonly used in the protest paradigm literature: framing and sourcing. Framing, or negative frames, refers to the ratio to the number of delegitimizing frames weighted by the number of words in each text. Sourcing, or quoted officials, refers to the number of times when NYT chose to quote mayors, police chiefs and other officials on the topic of protests also weighted by a text's word count. The values for both framing and sourcing were extracted using automated text analysis, the procedure of which is described in the Methods section in greater detail.

Two other variables were included as predictors in the model to control for potential confounding factors:

- learning: this control variable demonstrated how much time had passed since the first day of the protests. It was included as it can help to indicate whether media learns to change their narrative over time. The intuition here was that when the violence first broke out, the media could have reacted in a more negative way, than in general, for a number of reasons. It could be that the violence during the protest events was shocking, overwhelming, etc. so that it overshadowed the normal intention of a liberal media outlet to support the movement in their coverage. In a hypothetical situation, as time progresses, the emotions tend to subside, the movement engages in a peaceful dialogue to explain what happened and now that the media learns about these development, it might decrease their negativity.

- the number of protest events: as the violence variable is converted to the percentage ratio, it would make sense to add control for the number of protest events for each observation. There was a difference in the inferences which could be made about a day with two protest events and a day with twenty. Yet, if half of the protests on both those days were violent, percentage-wise it would be the same (both 50%). To make sure that the difference was taken into account, it was decided to include the number of protest events as a control variable.

### 3.3 Data Collection

This section is devoted to the choices and processes of collecting textual data from the New York Times, as unlike collecting data for the independent variables, the texts were not extracted from the pre-made dataset. Therefore, it required creative, but careful handpicking of text as data as well elaborate justification of the selection techniques. The key promise of this innovative way to organize data was that, if organized correctly, it contributes greatly to the investigation of empirical factors driving the tone of the media coverage to protests events. The data collection procedure consisted of three main components. First, I extracted the news coverage in text format from the New York Times archive. Next, I downloaded the ACLED's US Crisis Monitor database on political violence during Black Lives Matter protests in 2020 for the first independent variable, violence. For the second independent variable, popularity, I manually compiled the values indicating daily popular support for the Black Lives Matter movement from the CIVIQS poll website.

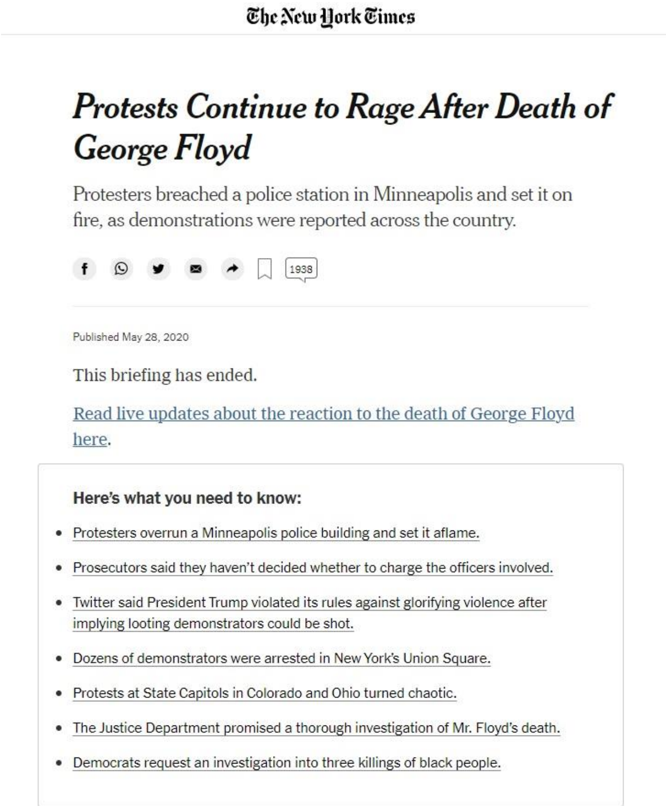
Access to the New York Times news archive (collecting articles since 1851) was made available by purchasing a digital subscription. As mentioned previously, the search was narrowed down to two weeks: May 26th to June 1st, and June 2nd to 8th. The search terms included "George Floyd Protests"; "Black Lives Matter". There were several other content sections in which the NYT publish their articles: most popular include "World"; "Business"; "New York", and others. For the sake of relevance and a fair balance of coverage, it was decided to narrow down the search to the "US" section only, which provided the coverage for events taking place nationwide.

The main criterion for an article to be included in the final dataset was to contain a text report on the protest activity. I excluded the articles that used the protests as a reference point to other topics, like discussing pre-election issues and candidates (e.g., "Trump's Looting and 'Shooting' Remarks Escalate Crisis in Minneapolis"). Thus, the actual number of gathered articles for the first week was 37; for week two the number was 33. This made 70 articles in total. Some of the articles were organized in a peculiar way: each could have contained several sub-articles that would generate a large briefing text on a particular topic. Thus, one article could have contained multiple reports on protest activity scattered all around the US. In the data collection for this research, those reports would all be collected and divided into separate text chunks depending on the state and date of the protest in the coverage.

After reviewing all the relevant articles for two weeks, the total number of such text chunks amounted to 192. That meant there were 192 reports on protest activity on a particular state at a particular date. As in the case of violence, the unit of analysis was transformed to be matching popularity on a state-date level. This meant that a text chunk contained coverage of protest activity in all cities within a certain state. For better clarity, I organized the units in question in the following screenshot attached to the screenshot:

Figure 3.2. The "hierarchy" of the media coverage units of the New York Times.

[Screenshot on the left made by the author. Retrieved on May 6th from https://www.nytimes.com/2020/05/28/us/george-floyd-national-guard.html.]


The New York Times

*Protests Continue to Rage After Death of George Floyd*

Protesters breached a police station in Minneapolis and set it on fire, as demonstrations were reported across the country.

1938

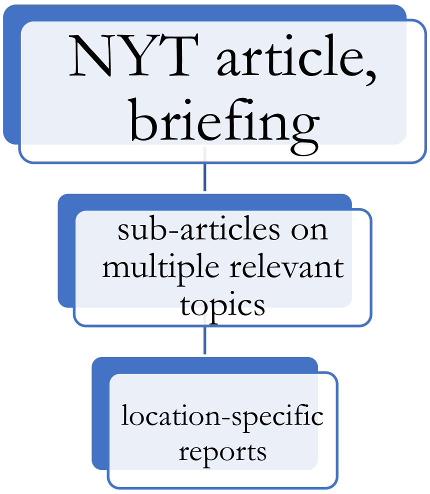
Published May 28, 2020

This briefing has ended.

Read live updates about the reaction to the death of George Floyd here.

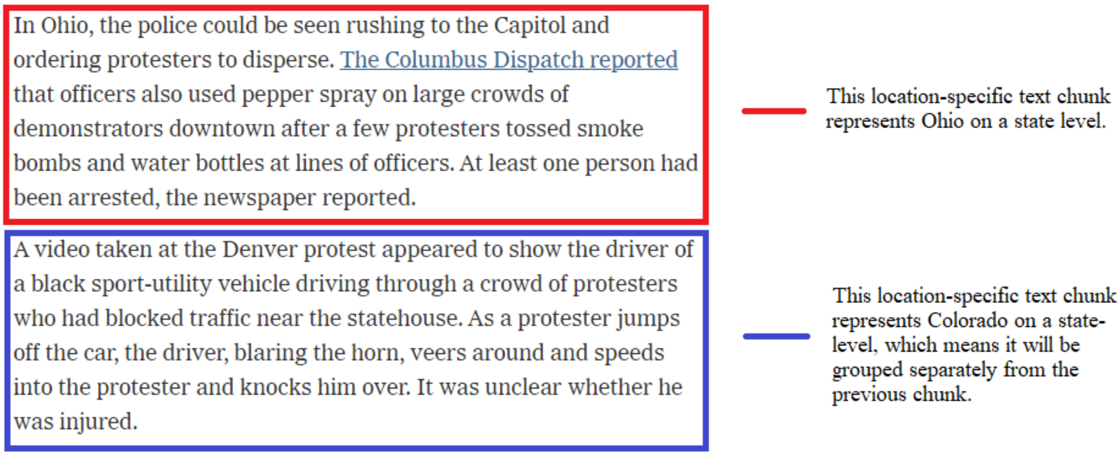
**Here's what you need to know:**

- Protesters overrun a Minneapolis police building and set it aflame.
- Prosecutors said they haven't decided whether to charge the officers involved.
- Twitter said President Trump violated its rules against glorifying violence after implying looting demonstrators could be shot.
- Dozens of demonstrators were arrested in New York's Union Square.
- Protests at State Capitols in Colorado and Ohio turned chaotic.
- The Justice Department promised a thorough investigation of Mr. Floyd's death.
- Democrats request an investigation into three killings of black people.




As can be seen in figure 3.2, some of the NYT articles were organized in the briefing form, which could include several sub-articles. On the screenshot on the left, the title "Protests Continue to Rage After Death of George Floyd" refers to the name of the article. The section "Here's what you need to know" represents a list of sub-articles. Figure 3.3 shows the content that NYT sub-articles consisted of: those were location-specific (city or town-wise) coverage text chunks. Such text chunks were compiled into an individual **state-date level text chunk**. For example, if multiple sub-articles coming from the same article contained the reports on protest activity from Los Angeles, San-Francisco, and Sacramento, they would be grouped together to a single text chunk, as on the state level they all represented California. Thus, if we were to locate it within the NYT hierarchy of media coverage units, it would take place between "sub-articles on multiple relevant topics" and "location-specific reports".

Figure 3.3 Two location-specific text chunks later coded as state-level chunks inside the same sub-article of the NYT reporting on George Floyd protests in 2020.


In Ohio, the police could be seen rushing to the Capitol and ordering protesters to disperse. The Columbus Dispatch reported that officers also used pepper spray on large crowds of demonstrators downtown after a few protesters tossed smoke bombs and water bottles at lines of officers. At least one person had been arrested, the newspaper reported.

A video taken at the Denver protest appeared to show the driver of a black sport-utility vehicle driving through a crowd of protesters who had blocked traffic near the statehouse. As a protester jumps off the car, the driver, blaring the horn, veers around and speeds into the protester and knocks him over. It was unclear whether he was injured.


## 3.4 Methods

This study employed multiple research methods. The ultimate purpose was to test whether there was a linear relationship between the predictor and response variables. This, however, required extracting the values for the response variable that consisted of two different measures: framing and sourcing. Quantifying framing and sourcing would allow demonstrating the number of times NYT chose to use a negative frame (framing) and quoted mayors, police chiefs, and other officials (sourcing). The tone of the coverage, or more specifically, the degree of its negativity was a response variable in regression model analyses. In these models, violence and popularity were the key predictors along with three control variables (fixed effects by state, protest events number, and learning). In order to transform text into numbers, I used dictionary-based automated text analysis.

### 3.4.1 Automated Text Analysis

The previous works on the protest paradigm established a dominant preference of using content analysis as their method of media coverage analysis. Usually, content analysis presupposes employing two independent human coders, the procedure of which requires both to pass a certain threshold (mostly >80%) of intercoder reliability. This would ensure that the coders interpreted the text in a similar manner, classifying them as per the pre-established set of coding rules. While human coding is a popular practice in qualitative research, which helps to determine the overall tone of the news articles, I decided to resort to quantitative text analysis methods for several reasons. The first reason is that the research goal in this thesis is not set around determining the number of the text chunks selected for analysis being either positive or negative. Rather, as mentioned earlier, the primary goal is to see how empirical factors, on average, might affect the variation in the use of framing and sourcing in each text. Additionally, the limited resources of this project would not have allowed for employing human coders nor would the research ethics allow me to do the coding, due to the potential bias. Automated text analysis was, therefore, one of the most appealing methodological alternatives for such a trade-off.

Automated text analysis (ATA) is a research method that came into use recently and was embraced with both enthusiasm and skepticism by the academic community. One of the main sources of concern, according to Grimmer and Stewart (2017, p. 2), to the complexity of the language: "automated content analysis methods will never the replace careful and close reading of texts". To overcome these limitations, researchers must come up with carefully crafted strategies that would reveal tendencies in writing that even a careful reader could miss. In the context of this study, I used ATA to gain a convincing insight into the tendencies in the journalistic practices that influence the tone of the media coverage, such as framing and sourcing. For this purpose, I employed a dictionary-based text analysis technique. According to

Grimmer and Stewart (2017), dictionary methods are supervised analysis methods that determine the frequency of the search terms specified in the lists known as dictionaries. In other words, the dictionary-based ATA would go through all the texts in the dataset, find the mentions were specified as the search terms, and provide a user with a number of how many search terms appeared in each text. Semantic context often would often get lost within such analysis. Nevertheless, Grimmer and Stewart (2017, p. 9) argued that dictionary methods are commonly used in analyzing tone, rather than the sentiment (which is structurally more complex and requires elaborate validation). Correctly specified terms in the dictionaries would contribute to obtaining an accurate frequency for both framing and sourcing, the count values of which would be used in regression analyses. The correctness of specifying terms in this research relied on the literature, theory, and data collection, and the validity of the research demanded its transparent discussion.

During the pre-reading stage of data collection, the media outlets had been observed to refer to authority officials whom they were asking and quoting on the topic of protests in a particular manner. If the media wanted to quote an individual holding a public position in the government-related office, then a media outlet would mention their position along with their last name with the use of direct or reported speech (e.g., said Mayor De Blasio). In this thesis, I refer to this group of quoted individuals as "officials": the number of quoted officials in the NYT coverage of George Floyd protests forms the sourcing proxy of the response variable in this study. The quoted officials were referred to by media with the use of unigrams (mayor, chief, spokeswoman, sheriff, and other). In regard to framing, the pre-reading stage data collection revealed that the predominant number of frames identifying the tone of the coverage of protests is also rather simplistic. The frames were predominantly composed of unigrams (single words) carrying a meaning of either support (e.g., protesters; peaceful; march; demonstration; and other) or condemnation and delegitimization (e.g., looters; dangerous; clashes; arrests, etcetera). For the sake of the research goals, I am interested in the NYT usage of the latter.

Kilgo and Harlow (2019) pointed out that negative media coverage focuses on four key frames when it comes to the delegitimization of progressive social movements. These frames are riot, confrontation, spectacle, and debate. In order to maintain the focus of negativity, the selection for the frames in the text analysis was narrowed down to riot and confrontation. This is motivated by several theoretical explanations. Firstly, the "debate" frame represented the media highlighting movements' true goals and demands (Kilgo and Harlow, 2019) and therefore showed signs of partial legitimization and recognition rather than otherwise. The "spectacle" frame, despite focusing on theatricality and dramatic effects of the protests (Kilgo and Harlow, 2019), was also omitted from the analysis, as there is no evidence that it could not potentially increase the support for the movement among the liberal readers. In the case of "spectacle" according to Kilgo and Harlow (2019, p. 518), the media "emphasizes the size of protests; […] odd aspects, like attire and clothing", the concept of which implied the ambiguity of the possible reactions by the readership. In the case of "riot" and "confrontation" frames, which were taken as the basis of forming text analysis dictionary, the possible reactions are much more limited to the negative:

- Riot: "The focus on the violence of protesters, on rioting, looting, description of the destructions, and chaos that protestors create in society" (Kilgo and Harlow, 2019, p. 518).
- Confrontation: "The focus on police versus protesters, [i.e.], arrests of protesters, conflicts with police, employment of police for enforcing protests" (Kilgo and Harlow, 2019, p. 518).

Both "riot" and "confrontation" frames are delegitimizing by definition: they view protesters as criminal elements of society that act as sources of violence, disruption, and unrest. The dictionary that was subsequently formed based on these framing directions included a set of the following search terms: [riot*, loot*, burn*, clash*, violent, chao*, disrupt*, destr*, vandal*, arrest*, skirmish*, brawl*, dangerous, unlawful, smash*, attack*]. In this dictionary, the (*) operator served a function of "umbrella" for all the words that had a similar "root". For example, if the dictionary consisted of only one search term (loot*), then the dictionary-based analysis would reveal how many times the words starting with loot- (loot, looter, looters, looting, looted, etc.) were mentioned across all the text chunks.

Specification of terms for the sourcing dictionary followed a similar logic based on the tendencies observed in literature and during the pre-reading stage. The list of the search terms included: [mayor, chief, sheriff, spokesman, spokeswoman, senator, deputy, secretary, governor]. The list included multiple authority officials of the local or state importance, which were observed to have a greater chance to be quoted on the protests, rather than use protests as a proxy to refer to something outside of the discussion (i.e., like then-president Trump).

After setting up the dictionaries for the search, the initial step for every text analysis method is preprocessing of the text. The texts were cleaned from the punctuation size, tokenized into unigrams and the dictionaries were transformed into document-feature-matrix (dfm). A document-feature-matrix is a matrix that contains each text chunk separately along with the number of specified dictionary terms mentioned in the text (Schoonvelde, 2020). There were two separate dfm-s that I used for this research: one for negative frames and one for sourcing officials. The total number and mean of negative frames, officials and protesters sourced across all 192 text chunks are shown in table 3.1 below.

Table 3.1. Results of dictionary-based analysis

| | № of texts with no mentions | № of texts with at least one mention | Total | Mean |
|---|---|---|---|---|
| **Negative Frames** | 88 | 104 | 429 | 2.234375 |
| **Officials Quoted** | 112 | 80 | 182 | 0.9479167 |

As can be seen in table 3.1, for both dictionary-feature-matrices (negative frames and quoted officials) in approximately half of the observations the analysis returned zero mentions. This had several important implications for the choice of the regression models. However, before discussing the distribution of the response variable, there was one more important step to be made. Since each text chunk had a varying number of words, it could have a different effect for every observation. For example, a chunk with 30 words in it and 2 negative frames might have had a profoundly more noticeable negative effect on the reader, than a chunk with 1000 words and 2 negative frames. Therefore, it was decided to create weighted values as per the word count in each cell. That way, it would reveal how many negative frames or quotations of officials/protesters each text chunk contained per 1000 words. The exact formula looked as follows:

**N/n * 1000**,

where **N** was the total number of mentioned search terms in question (negative frames, sourcing officials) in a text; **n** is the total number of words in a text. The **n** divided by the **N** and multiplied by a thousand will give a total number of the search terms in question per 1000 words of a text in each narrative. The distribution of the response variable is shown in figures 3.4 and 3.5 below for each document-feature-matrix.

Figure 3.4 The distribution of negative frames used by the New York Times in their coverage of George Floyd Protests weighted by the number of words in each article.

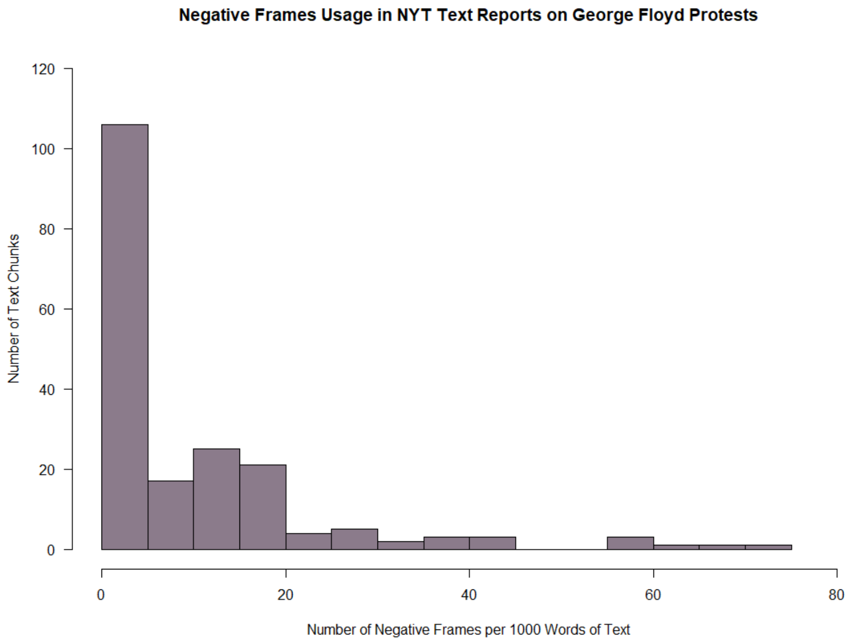
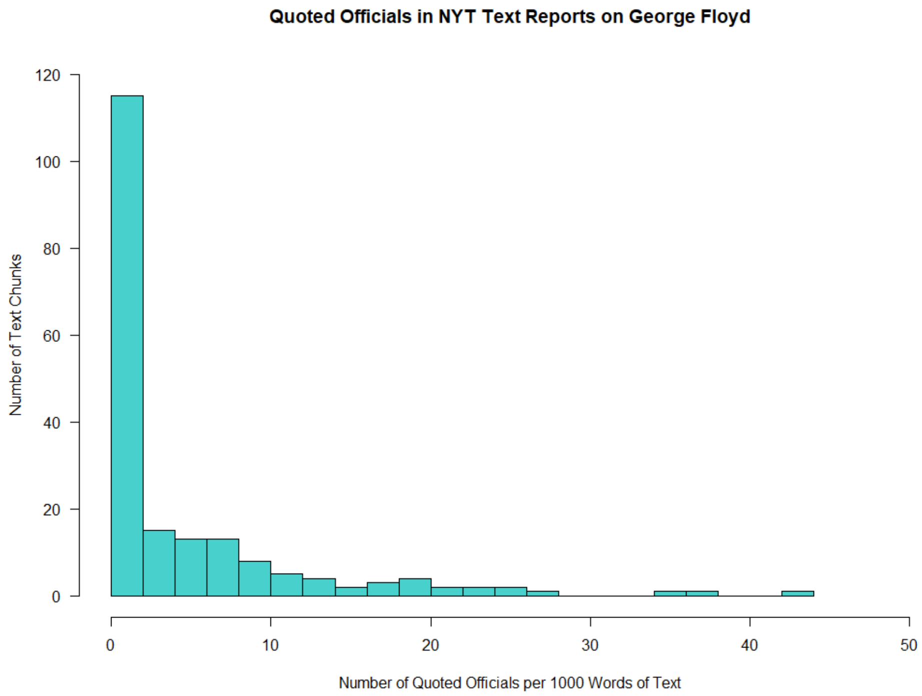


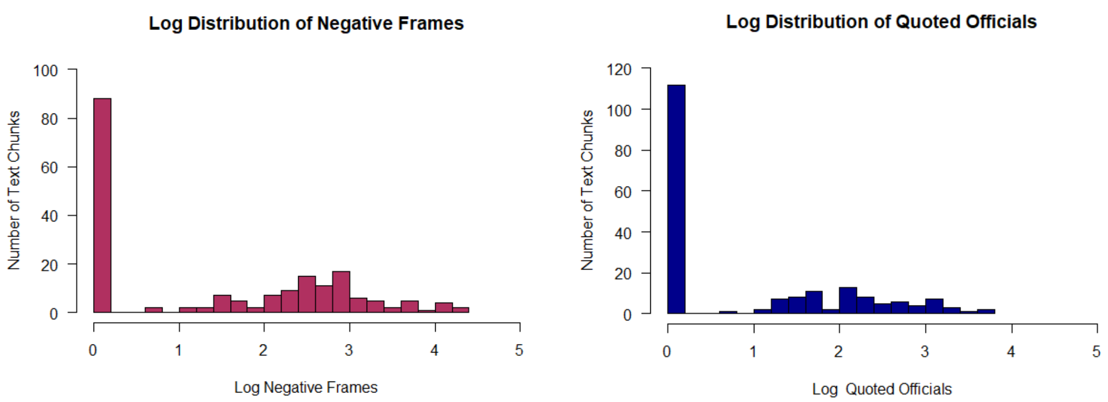

Figure 3.5. The distribution of officials quoted by the New York Times in their coverage of George Floyd Protests weighted by the number of words in each article.



### 3.4.2 Model Selection Procedure and Justification

The histograms in figures 3.4 and 3.5 show a similar issue of the response variable values skewed to right. In other words, in both response proxies (negative frames, sourcing officials) approximately half of the observations returned zeroes. This meant that the automated text analysis did not find any of the pre-established search terms in the text. Assuming that the search terms were valid, I decided to move on to solving the issue of skewed distribution through the logarithmic transformation. Log-transformation is a commonly used technique when it comes to dealing with highly skewed distributions. According to Benoit (2011), the log-transformation is a convenient method of transforming variables into values that follow "log-normal distribution", the untransformed form of which would remain otherwise skewed. The formula that I used for log-transformation looked as follows, with **R** referring to the response proxy in question (either framing or sourcing): log ((**R**)+1). Unfortunately, logarithmic transformation did not yield the desired results. The distribution of both response proxies (negative frames and sourcing officials) remained highly skewed, as figure 3.6 shows.

Figure 3.6. Logarithmic distribution of the response variable proxies (negative frames in maroon, quoted officials in blue).



Given that the log-transformation did not resolve the problem of skewed distribution in the response variable, it was decided to navigate through various regression models that address the problem of excess zeroes. Consequently, trying out Poisson, negative binomial, zero-inflated negative binomial regression did not yield a desirable result due to the reasons specified in **Appendix A**. The potential resolution to this issue demanded a better understanding of the reasons behind the excess zeroes in the dataset. After trying out several different alternatives, I decided to use two separate regression models for each response proxy: logit regression and OLS regression.

- For the logit models, I re-coded the response variable from continuous (from 0 to + infinity) to binary (from 0 to 1). The logit model of that design was expected to predict the choice of the NYT to be negative (1) in their coverage (either by using negative frames or quoting officials) or not (0). To put it simply, the logit models were supposed to answer the question: what are the odds of New York Times being negative or not in their coverage of the first two weeks of George Floyd protests in 2020?
- For the OLS regression models, I omitted all the zero values from both response proxies for one specific purpose. Keeping values that are bigger than 0 in the logit model meant that the NYT chose to be negative in their coverage of George Floyd protests. Given such choice, the OLS models were expected to predict the variation in the degree of negativity of the NYT through the numbers of negative frames and quoted officials. In other words, the OLS regression was supposed to provide an answer to the following question: If the New York Times chooses to be negative in their coverage, then to what degree is it negative?

Such decision to split theorized model in two types of models took place in the final stages of the research, after the theory-building. For transparency reasons, I decided to slightly modify the initially stated hypotheses and state new ones in this section, instead of rewriting and presenting them as originally planned. Since I had two models to work with, logit and OLS, the number of hypotheses doubled. The changes in hypotheses did not affect the relationship between the predictors and the response variable. Instead, they were adapted to better match the nature of their respective models. New hypotheses are listed below in the corresponding order.

**Logit Models:**

- H1a: An increase in violence causes an increase in the odds of the NYT using negative frames in their coverage.
- H1b: An increase in violence causes an increase in the odds of the NYT quoting officials in their coverage.
- H2a: An increase in popularity causes a decrease in the odds of the NYT using negative frames in their coverage.
- H2b: An increase in popularity causes a decrease in the odds of the NYT quoting officials in their coverage.

**OLS Regression Models:**

- H3a: If the NYT decided to be negative in their coverage, an increase in violence causes an increase in the number of negative frames.
- H3b: If the NYT decided to be negative in their coverage, an increase in violence causes an increase in the number of quoted officials.
- H4a: If the NYT decided to be negative in their coverage, an increase in popularity causes a decrease in the number of negative frames.
- H4b: If the NYT decided to be negative in their coverage, an increase in popularity causes a decrease in the number of quoted officials.

Before turning to the findings section, it is necessary to mention several other important data transformations took place. Firstly, I had to remove all observations bound to the DC state (22 cells). The data on popularity from CIVIQs did not include the District of Columbia as a separate state, which created missing rows when applied to the model. I have also removed the text chunks that were attached to the last two days because those contained only 4 observations.

The abnormally low number of observations in those two days affected the results in a significant way, especially when looking at the effect of learning as the control variable. For the same reason, I removed 5 outliers in the violence variable, which had a 100% percentage of violent events, but given the low amount of protest events in total (1 or 2). This resulted in the final sample of 163 observations.

Finally, in OLS regression models, the response variable values in both framing and sourcing were subject to the log-transformation due to the skewed distribution of some of its values. After the log-transformation, the distribution for both negative frames and quoted officials became much closer to normal, as can be seen in figures 3.7 and 3.8. This meant that the coefficients would have to be exponentiated during the interpretation of the model's summary statistics. One of the predictor variables, the percentage of violent protests or violence, was also subject to the log-transformation due to the same reason as framing and sourcing. Although it did not fully resolve the skewed distribution problem for violence due to the fact that the predictor contained zero values, unlike response proxies, the log-transformed values were still used for the OLS regression models.

Figure 3.7. The distribution of texts with at least one negative frame before (left) and after (right) log-transformation.

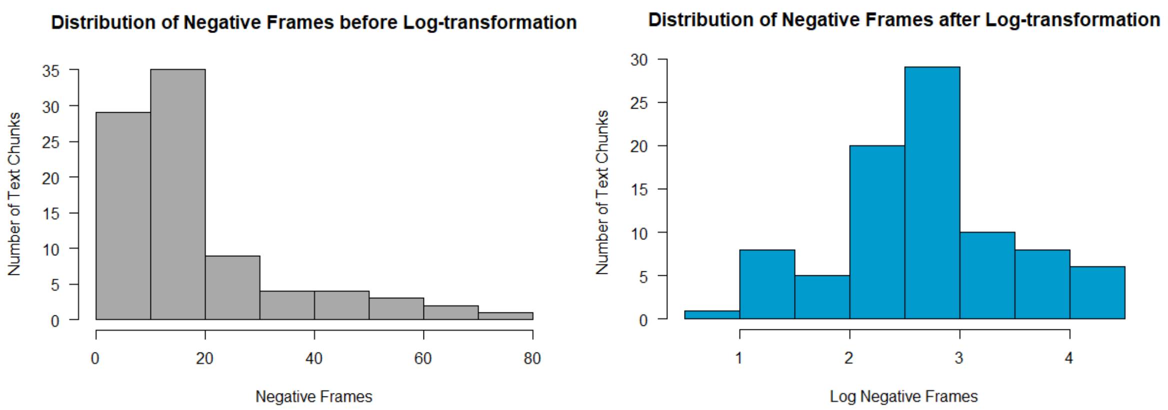
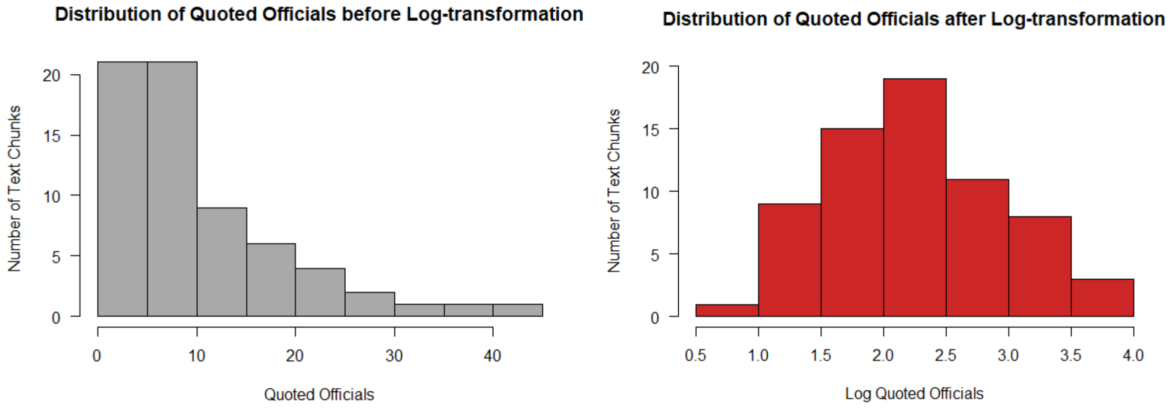


Figure 3.8. The distribution of texts with at least one quoted official before (left) and after (right) log-transformation.



# Chapter 4: Results and Discussion

This thesis chapter provides the overview and interpretation of the findings. The first part is devoted to the discussion of the results yielded by logit and OLS regression models. The models analyzed the impact of violence and popularity with control variables (number of protest events and learning) on the negativity of the media coverage (framing and sourcing). In the second part of the chapter, I discuss the potential confounding factors and other limitations of this research. Lastly, I highlight the implications for future research as well as the overall contribution of this thesis.

## 4.1 Logit Models

### 4.1.1 Logit Model: Framing

The findings of the logit model of negative frames demonstrated a rather counter-intuitive trend, where learning turned out to be a better predictor than violence and popularity when it comes to framing (table 4.1).

Table 4.1. Summary statistics of logit model for negative Frames

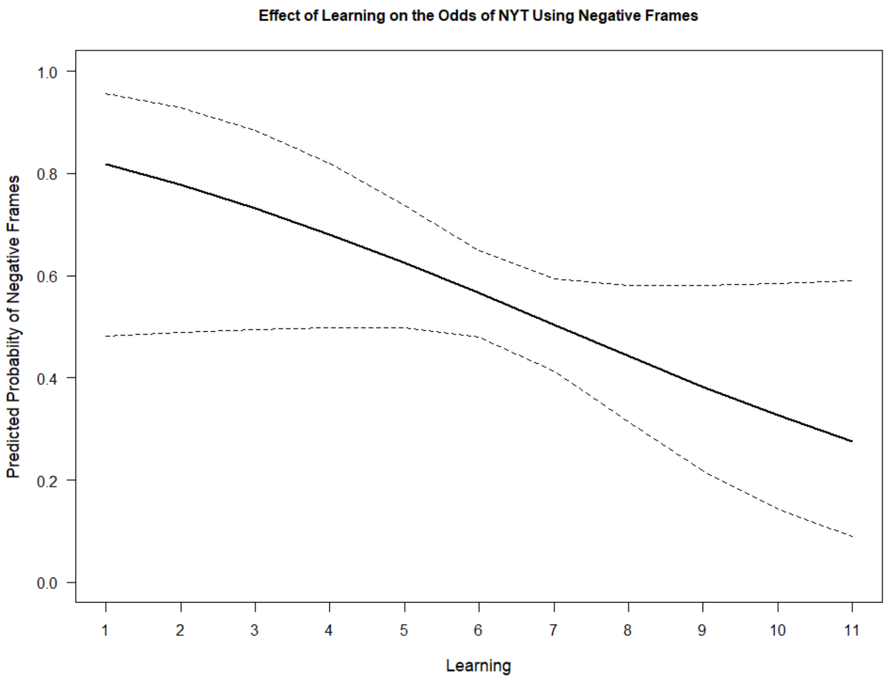
| **Variables** | **Coefficients** | **Std. Error** | **Exp.** | **z-value** | **Pr(>\|z\|)** |
|---|---|---|---|---|---|
| (Intercept) | 0.958551 | 4.911771 | 2.6079155 | 0.195 | 0.8453 |
| Support | 0.027309 | 0.041099 | 1.0276856 | 0.664 | 0.5064 |
| Neutral | -0.044702 | 0.133370 | 0.9562827 | -0.335 | 0.7375 |
| № of Events | 0.001561 | 0.012328 | 1.0015622 | 0.127 | 0.8992 |
| Learning | -0.246813 | 0.144751 | 0.7812871 | -1.705 | 0.0882 . |
| Violence | 0.016706 | 0.020882 | 1.0168463 | 0.800 | 0.4237 |

Significance codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Null deviance: 225.22 on 162 degrees of freedom

Residual deviance: 207.50 on 157 degrees of freedom
AIC: 219.5 Number of Fisher Scoring iterations: 4

In fact, learning was the only statistically significant variable in the logit model, according to the p-value of 0.0882. After exponentiating the log coefficient of the learning variable, the returned value, or odds ratio, was equal to ~0.74. This meant that with every single increase in learning units, the odds of the NYT writing a state-date level report with at least one negative frame decreased by 0.26. The overall negative effect of learning on the predicted probability of the New York Times using negative frames in their coverage can be seen in Figure 4.1 along with the upper and lower bound confidence intervals indicated by two dashed lines. Such finding can be explained by a similar intuition, in which I justified the inclusion of learning as a control variable. While it seemed that violence and popularity (support and neutral variables) were not statistically significant, it might have been that over time, the NYT changed or rather, adapted their views on the protests. This resulted in the decreased number of negative frames over time, irrespective of violence during protests or the movement's popularity.

Figure 4.1. The effect of learning on the odds of the NYT using negative frames in their coverage of the first two weeks of George Floyd protests in 2020.



#### 4.1.2 Logit Model: Sourcing

The logit regression model of sourcing officials presented the patterns that for the most part corresponded to theoretical expectations. Table 4.2 demonstrates the results for the logit model for NYT choice of whether to quote an authority official or not in their reports on George Floyd protests.

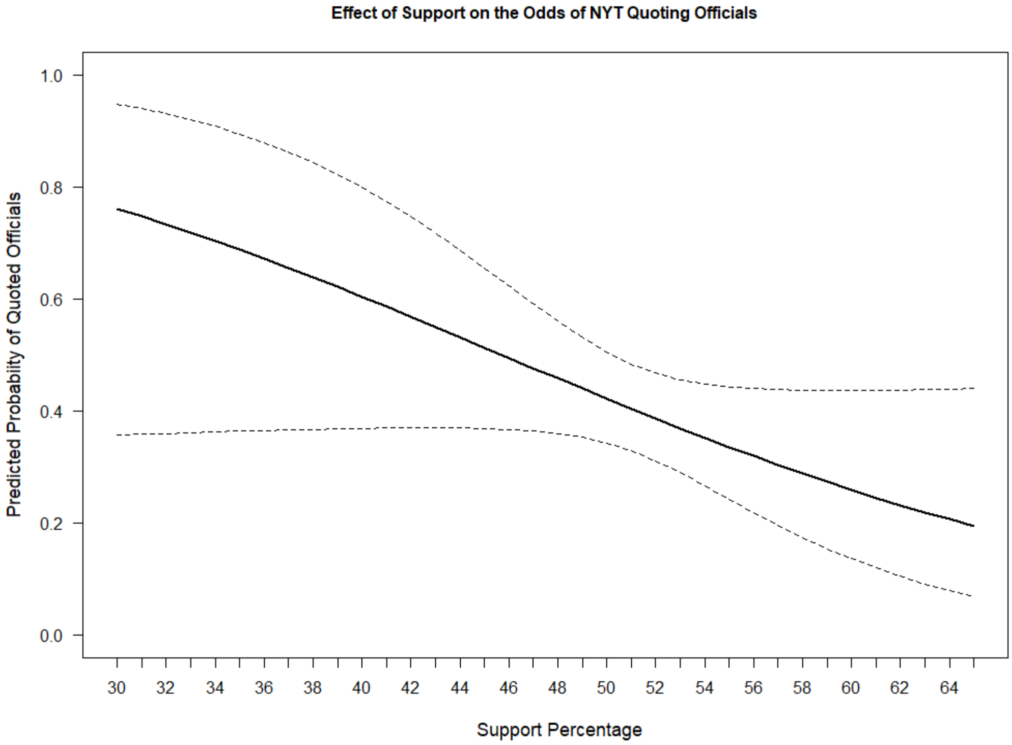
Table 4.2. Summary statistics of logit model for quoted officials.

| Variables | Coefficients | Std. Error | Exp. | z-value | Pr(>\|z\|) |
|---|---|---|---|---|---|
| (Intercept) | 7.399392 | 4.994256 | 1634.9904369 | 1.482 | 0.1385 |
| Support | -0.073715 | 0.041614 | 0.9289365 | -1.771 | 0.0765 . |
| Neutral | -0.141905 | 0.134324 | 0.8677034 | -1.056 | 0.2908 |
| № of Events | 0.022348 | 0.012482 | 1.0225994 | 1.790 | 0.0734 . |
| Learning | -0.279950 | 0.146040 | 0.7558217 | -1.917 | 0.0552 . |
| Violence | -0.002249 | 0.018775 | 0.9977530 | -0.120 | 0.9046 |

Significance codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Null deviance: 220.03 on 162 degrees of freedom
Residual deviance: 210.44 on 157 degrees of freedom
AIC: 222.44 Number of Fisher Scoring iterations: 4

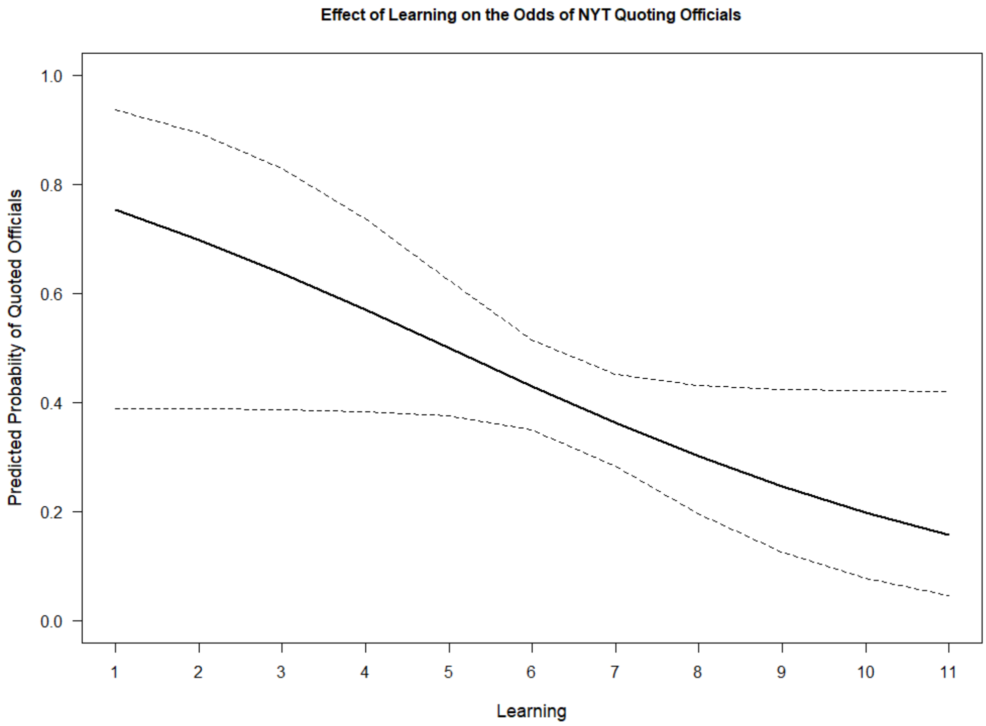
While the violence during the protests deemed itself statistically insignificant as a predictor variable, the popular support for the movement seemed to have a slight adverse effect on the chances of quoting officials in the NYT reports on George Floyd protests. Figure 4.2 below shows the effect of the US citizens' support on the predicted probability of the New York Times quoting officials in their coverage. The exponentiated coefficient for the support variable equaled to ~0.92. This meant that with a single unit increase in the percentage of people expressing support for BLM, the odds of NYT quoting an authority official decreased by 0.08. This finding corresponded to theoretical expectation that the NYT would slightly adjust to the rising popularity of the movement and quote less officials, who are not related to the movement in any ways.

Figure 4.2. The effect of support of the US citizens for BLM on the odds of the NYT quoting officials in their coverage of the first two weeks of George Floyd protests in 2020.




The number of events demonstrated a slight positive effect (increase by 0.02) on the NYT choice of whether to source an official or not. This makes general theoretical sense, in which the media would be expected to turn to the authorities to ask them to react on the protests, given that the number of protests would increase. Lastly, the learning variable once again was a statistically significant predictor with a solid negative effect on the response variable once again with the exponentiated coefficient of ~0.75. Such a result can be interpreted in the same way as in the case of negative frames: with one unit increase in days passed since the first day of protests, the NYT's odds of quoting officials decreased by 0.25. The overall effect of learning on the odds of the NYT to quote an authority official is shown in figure 4.3 below. In general, these findings are rather surprising in the sense that the factor of time passed had a more profound effect on the response variable than the empirical factors, such as violence and popularity.

Figure 4.3. The effect of learning on the odds of NYT quoting officials in their coverage of the first two weeks of George Floyd protests in 2020.




The model fit test for logit models was decided in favor of Tjur's R-squared statistics. Initially, the model fit was checked with Hosmer and Lemershow's goodness of fit test, which was a commonly used method of model fit assessment. The test yielded p-values of 0.2 for negative frames and 0.55 for sourcing officials which spoke of a convincing evidence against the null hypothesis. Yet, according to Allison (2014), Hosmer-Lemershow's test contained numerous theoretical drawbacks, mainly based on the arbitrary dependence on the number of groups. Tjur's R-squared is a relatively recent, yet already established goodness-of-fit method for the models with binary outcomes (Allison, 2014). It uses the coefficient of discrimination instead of the coefficient of determination, which according to Tjur (2009, p. 366) is easier to calculate and interpret, but at the same time is equivalent in the accuracy of the "measure of explanatory power". Tjur's R-squared statistics yielded the following result for each of the logit regression models:

- Negative Frames Model: Tjur's R2: 0.104
- Sourcing Officials Model: Tjur's R2: 0.059

## 4.2 OLS Regression Models

The Ordinary Least Squares regression models for framing and sourcing were designed to predict the degree of negativity via variation in the log-transformed values of negative frames and quoted officials. The OLS models' findings revealed several important implications for the research. While the OLS model for negative frames yielded the results, which were only in part consistent with theoretical expectations, the OLS for quoted officials showed that all the model's predictors were statistically insignificant. These and other several factors hinted at potential problems with the sample size and model, which I discuss in detail in the next section of this chapter.

### 4.2.1 OLS Regression Model: Framing

The summary statistics available in table 4.3 below shows the coefficients of statistically significant predictors in the negative frames OLS model.

Table 4.3. Summary statistics of OLS regression model for negative frames.

| **Variables** | **Coefficients** | **Std. Error** | **Exp.** | **t-value** | **Pr(>\|t\|)** |
|---|---|---|---|---|---|
| (Intercept) | -4.372643 | 2.900705 | 0.01261784 | -1.507 | 0.1356 |
| Support | 0.043388 | 0.025073 | 1.04434250 | 1.730 | 0.0874 . |
| Neutral | 0.158191 | 0.076170 | 1.17138978 | 2.077 | 0.0410 * |
| № of Events | -0.005506 | 0.006363 | 0.99450881 | -0.865 | 0.3894 |
| Learning | 0.232087 | 0.098036 | 1.26122926 | 2.367 | 0.0203 * |
| Violence | 0.269941 | 0.120947 | 1.30988682 | 2.232 | 0.0284 * |

Significance codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.7488 on 81 degrees of freedom
Multiple R-squared: 0.1313, Adjusted R-squared: 0.07764
F-statistic: 2.448 on 5 and 81 DF, p-value: 0.04062

Since the response variable was log-transformed, there was a need to exponentiate the coefficients of each statistically significant predictor (except violence). All of them have demonstrated a positive effect on the percentage of negative frames after exponentiation. In terms of popularity, the proxy Neutral, which stood for people treating Black Lives Matter in a neutral way seemed to have a more profound effect on the response variable than popularity. Figures 4.4 and 4.5 below show the effects of log support (4.4) and neutral attitudes (4.5) percentages on the predicted log negative frame values plotted along with corresponding confidence intervals. It is important to mention that plots for OLS regression model on negative frames show the effect of statistically significant predictors in log values of negative frames due to the previous log-transformation of the response variable. The plots of exponentiated effect for all statistically significant predictors along with justification of using plots with log-transformed values can be found in Appendix B.

Once the coefficient is exponentiated, it shows that a single unit increase in the percentage of people with neutral attitudes towards the BLM movement increased the degree of negativity in framing by 17%. This could be due to the fact that some of the people expressing support initially shifted in their opinion from supportive to the moderate neutral (possibly due to the violence that took place), therefore making the prospects of supporting Black Lives Matter less appealing.

Figure 4.4. The effect of support on predicted log number of negative frames in the NYT coverage of the first two weeks of George Floyd protests in 2020.

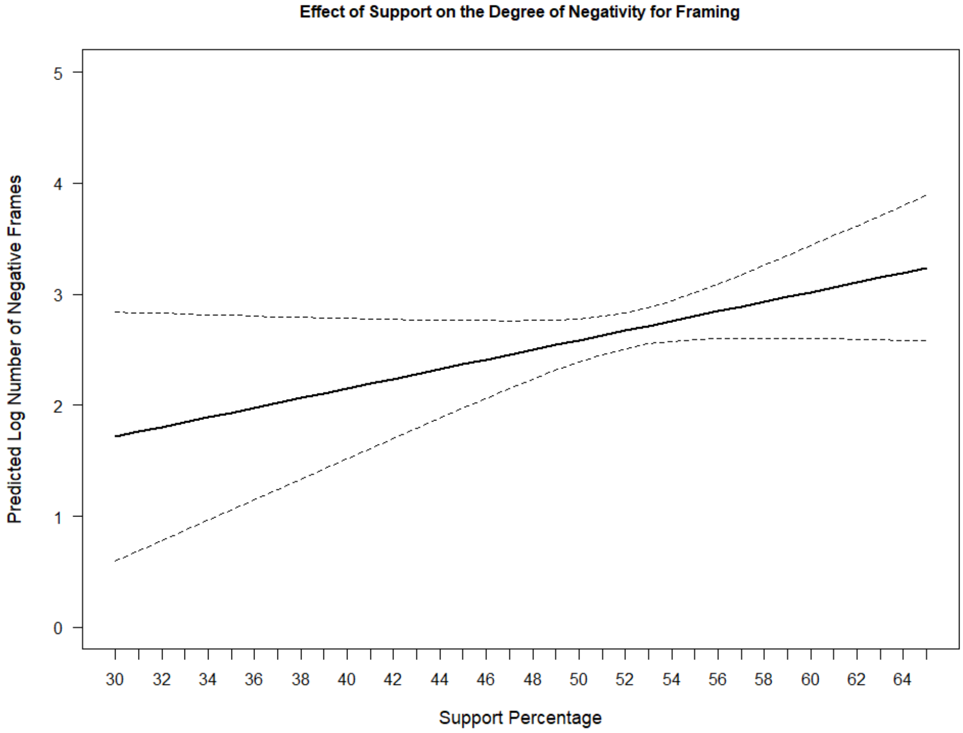
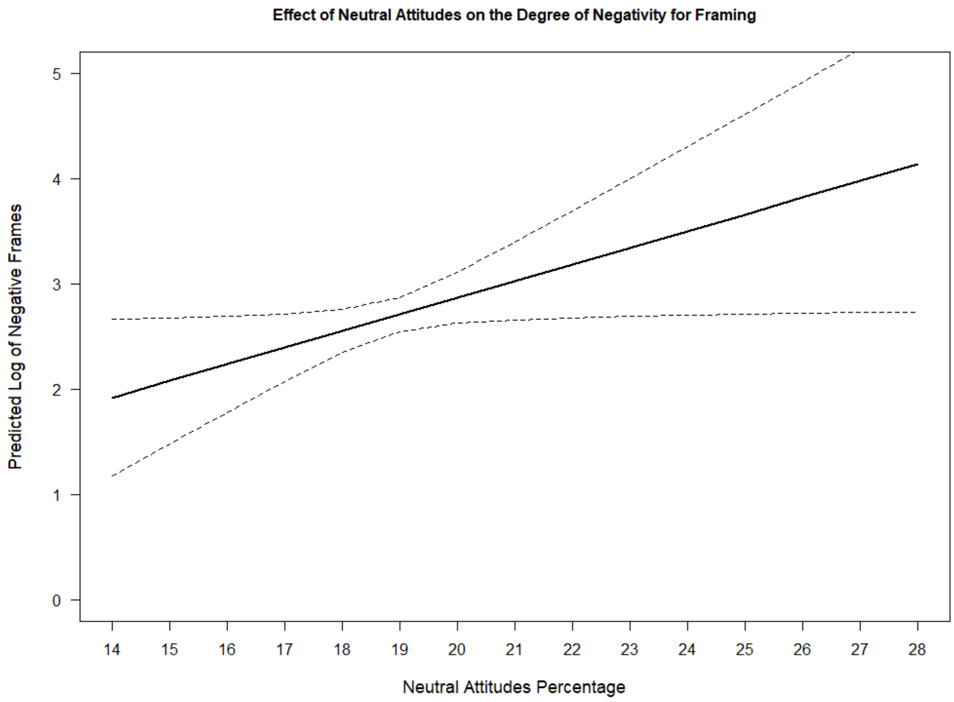
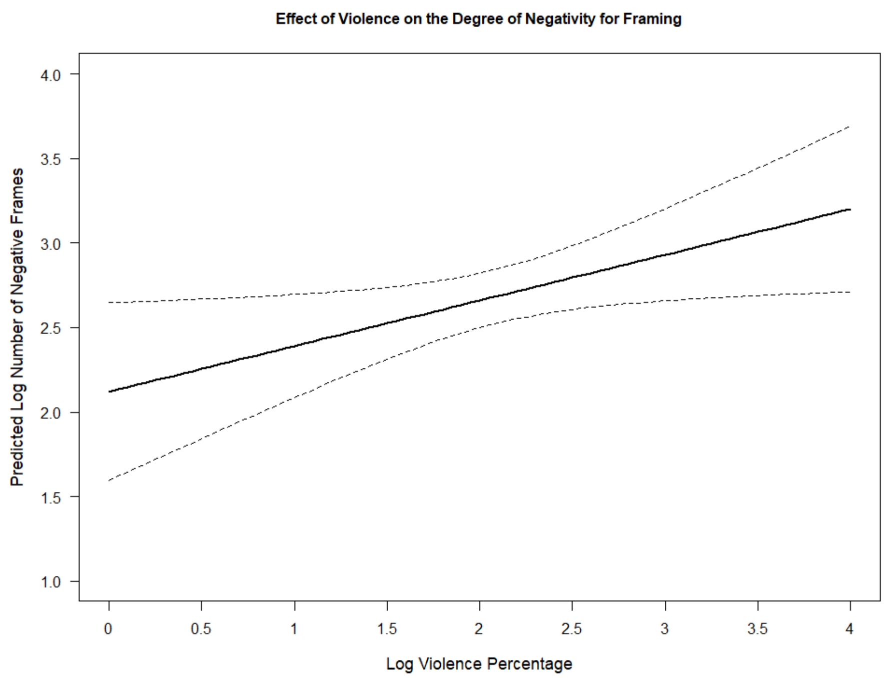



Figure 4.5. The effect of neutral attitudes towards BLM on predicted log number of negative frames in the NYT coverage of the first two weeks of George Floyd protests in 2020.



The other independent variable violence has shown a slight positive effect on the degree of negativity of the NYT. Figure 4.6 demonstrates the positive effect of log percentage of violent events on the log number of negative frames as well as the confidence intervals. However, since the violence variable was log-transformed as well, the rules of coefficient interpretation were different. According to UCLA Institute for Digital Research and Education, if both predictor and response variables were log-transformed, the effect could only be interpreted through a joint change in percentage. That is, there was no need for exponentiation in the case of the violence variable because it was log-transformed prior to running the OLS model. In order to calculate the change in percentage for the response variable, given the constant change in the predictor variable (for example, 10%), I put the value of 1.10 (the 10% change) and put it to the power of coefficient of violence (.269941). 1.10^.269941 equaled to 1.026062, which meant that for any 10% change in violence, the degree of negativity in framing increased by 2.6% (UCLA). As in the case with support and neutral attitudes towards BLM, the plot for exponentiated effect of violence on the number of negative frames will be attached to Appendix B.

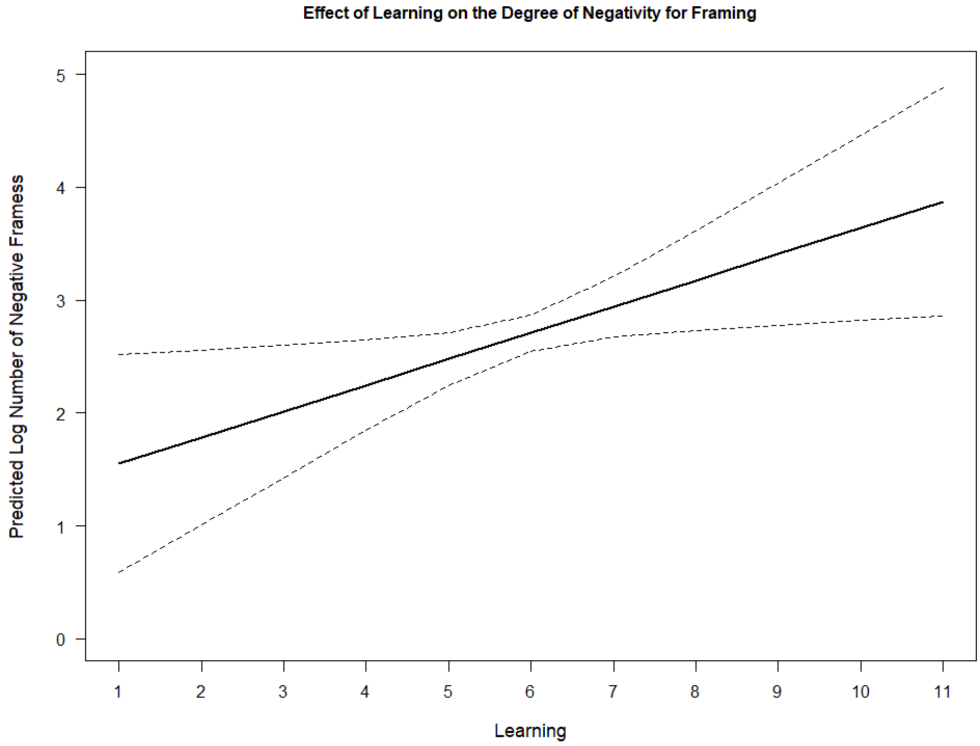
Figure 4.6. The effect of log violence on predicted log number of negative frames in their coverage of the first two weeks of George Floyd protests in 2020.



This finding, however, was a subject to a more cautious interpretation. Since each observation was organized by state and date and not by article, there was a possibility of some observations being dependent on one another. For example, they would exhibit mutual dependency in the case coming from the same article, but different sub-articles, or if both would come from the different articles, but taking place on the same date and report the protests from the same state. In order to overcome such complexities, as part of the robustness check procedure, I calculated clustered standard errors for all models clustering them by state (full summary statistics for all models can be found in Appendix C). This did not introduce any significant corrections to the coefficients of summary statistics I already described, except for the violence variable in the OLS regression model of framing. After applying the cluster by a state to the linear regression, the statistical significance of the violence variable dropped considerably (p-value increased from 0.0284 to 0.13).

Finally, learning, or the number of days passed since the first day of the protest, also seemed to have a profound positive effect on the negativity of framing. The overall effect of learning on the log number of negative frames as well as the confidence intervals marked with dashed lines is shown in figure 4.7 below. Exponentiated coefficient shows that with one unit increase in the learning variable, the degree of negativity in framing increased by 26% (see Appendix B for the plot with exponentiated effect). This was a rather counter-intuitive discovery, given the results of the logit models of negative frames and quoted officials, in which learning actually decreased the odds of NYT being negative in their coverage of George Floyd protests. Such contradicting trend could mean that while the learning tended to decrease the odds of NYT having any negative frames in their reports, such restraint could become an amplifying factor in the case of negative frames, once NYT decided to be negative in their coverage.

Figure 4.7 The effect of learning on predicted log number of negative frames in their coverage of the first two weeks of George Floyd protests in 2020.




### 4.2.2 OLS Regression Model: Sourcing

Regarding the OLS regression on quoted officials, the results did not show that the model had any statistically significant predictors whatsoever. The summary statistics is shown in table 4.4 below.

Table 4.4. Summary statistics of OLS regression model for quoted officials.

| Variables | Coefficients | Std. Error | Exp. | t-value | Pr(>\|t\|) |
|---|---|---|---|---|---|
| (Intercept) | 4.495412 | 3.354192 | 76.2193399 | 1.340 | 0.185 |
| Support | -0.026358 | 0.024207 | 0.9741922 | -1.089 | 0.281 |
| Neutral | -0.055982 | 0.090165 | 0.9488172 | -0.621 | 0.537 |
| № of Events | -0.001060 | 0.006531 | 0.9988536 | -0.162 | 0.872 |
| Learning | 0.036097 | 0.106383 | 1.0468194 | 0.339 | 0.736 |

| | | | | | |
|---|---|---|---|---|---|
| Violence | -0.020969 | 0.118560 | 0.9992460 | -0.177 | 0.860 |

Significance codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.6832 on 60 degrees of freedom
Multiple R-squared: 0.1086, Adjusted R-squared: 0.03435
F-statistic: 1.462 on 5 and 60 DF, p-value: 0.2156

As can be seen, the predictor variables in the model not only deemed statistically insignificant, according to the results but also show a very modest effect on the response variable. These findings hint at several possible reasons, one of which is the absence of a linear relationship between the predictors and response variable. One other possible inference is that the overall sample was too small (163 observations) to explain any variation of the negativity through sourcing officials.

### 4.3 Limitations and Confounding Factors

The overall results demonstrate that despite the partial success of logit and OLS (in the case of negative frames) models explaining the choice of NYT to be negative or not (and if yes, to what extent) in their coverage, both regression analyses pointed at the specific limitations of this research. Additionally, despite the fact that the logit models' results generated some interesting implications, a check of models' quality with observed and expected values yielded ~64% of the correct predictions for framing and ~67% for sourcing. These percentages speak of the model predicting the odds of NYT being negative better than the random probability of a binary outcome (50%) but still is far from the ideal. Additionally, the adjusted R-squared values of the OLS regression models on both framing and sourcing imply that the model is overfitted. In this part of the discussion section, I go over the potential confounders and other limitations that could have affected the models' results.

Police brutality was one possible uncontrolled confounder that could have affected the negativity of the coverage and some of the predictors at the same time. Despite the fact that the protests could be violent, the NYT could react to the protests more sympathetically, if there was a solid presence of police brutality against the protesters. Given the theoretical expectations on matching ideologies with the protesters (Weaver and Scacco, 2013), it would be reasonable to expect that in such a scenario, NYT would be more supportive towards BLM, which would, in turn, increase the popularity of the movement even further. Indeed, as observed during the pre-reading stage in some of the articles, there were several reports on attacks on the peaceful protesters and the journalists. For this, however, to be quantified and included in the analysis, the research would require data on the police brutality occasions. While there were several non-profit organizations that reported on the use of excessive force during the protests one of the most promising looking ones was the project called Mapping Police Violence (mappingpoliceviolence.com). Nevertheless, despite being organized in a very detailed and professional manner, allowing the data to be used in secondary research, the dataset of the project contained only the records of individuals killed by the police. There were, however, other forms of police violence, such as using pepper spray, riot gear, and other means of violence that did not lead to lethal outcomes. The other sources, like ProPublica (projects.propublica.org/protest-police-videos/), were heavily dependent on the documented evidence of the witnesses. This meant that their dataset was too scarce to provide the full picture of events (68 documented incidents) during the first weeks of George Floyd protests. Thus, for transparency reasons, it was necessary to concede that there was a lack of alternatives for measuring police brutality, which was a significant confounder that could have affected the results of both logit and OLS models.

The other unaccounted confounding factor, which would be extremely hard to quantify at best, is to track the personal biases of the journalists, editors, and other staff responsible for the rhetorical choices of the coverage. While it might be that as an institution, NYT can be regarded as a left-leaning US media outlet with progressive values, journalists and other writing staff may not share views on certain tactics of the social protests. McLeod and Detenber (2006,

p. 4) argue that the reasons for media engaging in protest paradigm range from "the biases of individual journalists" to "organizational imperatives and sociocultural worldviews". The works on the protest paradigm, however, do not discuss the ways in which it would be possible to measure those individual biases, or track them, at all.

As far as the other limitations go, a lot of them were related to the data collection and organization procedures. The main concern was the organization of observations on the state-date level. Despite making theoretical sense, in part, it was also a decision that had to be made given how NYT articles and some of the predictors were organized. Of course, if the NYT news articles were organized in the way that a single article would contain a single report on a protest activity, it would be easier to just collect the articles in their original form. But since one article from New York Times might consist of multiple sub-articles, which in turn might contain multiple location-specific (city or town-wise) reports, the data organization required more creativity and involved an active pre-reading stage. On the other hand, if this thesis dataset collected only location-specific reports, then the cumulative effect of violence would practically be nullified. Organizing data by location-specific observations would equate a single text chunk to be reporting on a single protest. If so, then the violence variable would range between 1 (violent protest) and 0 (non-violent protest). In this research, I was more interested in the cumulative effect of violence on the state-date level but organizing data in the way described above might prove more efficient for future researchers.

Another important concern was dedicated to the validity of negativity measures, namely sourcing. As mentioned in the methodology chapter, the key limitation of using the dictionary methods as a tool of extracting meaning from text is that such text analysis often fails to get behind the context. Thus, the values obtained from the text on the number of quoted officials ignore in which context the officials were quoted: some of them might could the resilience of protesters, some could in fact label activists as looters, others could send a mixed message. My guidance in that regard was that during the pre-reading stage, I observed the last two groups appear in the text far more often than those who would express unequivocal support for the George Floyd protests. Apart from that, in some situations, the readers might expect the media to quote the authorities on many relevant occasions. For example, if the peaceful protesters could have been hurt by the excessive force of policemen, the public would most likely demand a commentary from the chief of the police. This limits the perceived freedom of choice of a media outlet and could have affected the model results in a detrimental way. Fortunately, the exact number of quoted officials was not important for this project per se, as was its potential to explain its variation through a linear relationship with the predictor variables.

Finally, if given more time and resources, this research could make use of contacting the NYT editorial board and journalists in order to gain more insights into the publishing practices of the media outlet. This study held the assumption that the content is published promptly and not edited after publishing. Since the time period of this study at the moment of writing analyzes the media coverage from almost a year ago, it would be reasonable to inquire to see the earlier versions of the reports on protests, if any. That way, even if the media indeed learns over time not to engage in protest paradigm against Black Lives Matter, the overall effects could have been different. And even though it is more than possible that the NYT would not be interested in revealing that kind of information, I believe that even the documented refusal would still benefit the future research in the same direction.



### 4.4 Discussion and Concluding Remarks

Despite the problems with the model, this research ends on a promising note. Testing the effects of empirical factors on the tone of the coverage was an innovative inquiry in its own way: the studies on protest paradigm in the past had pursued the exploratory research objective for the most part. One of the key findings of this study was that learning, or time passed since the first day of protests, decreases the odds of the NYT engaging in the use of negative frames as well as sourcing officials. Apart from that, the results of the OLS model on framing suggest that if the NYT decides to be negative in their coverage, then learning amplifies the degree of negativity by 26%. This was a rather contradictory and therefore surprising discovery. Overall, this finding partially confirms Gottlieb's (2015) claim that media tend to change the focus of their framing over time. This thesis does not reveal that media becomes necessarily more supportive of the movement, but rather less critical after time passes irrespective of the violence during protests and the movement's popularity.

In regard to the latter two, the independent variables of this research demonstrated their relevance in only two models out of four. Violence deemed itself statistically significant only in the linear regression model on framing. This did not practically mean that the NYT is never aware of the violence levels when deciding to be negative through framing or sourcing, but rather that violence played a much less significant role for the negativity of the coverage than expected. Nevertheless, it showed that it has a slight increase in the degree of negativity of framing: with a 10% change in the predictor, the negative frames increase by 2.6%.

Popularity was statistically significant in two models: logit model for quoted officials and OLS regression for negative frames. In regard to US citizens feeling neutral towards BLM, the neutral attitudes, like violence, demonstrated statistical significance only in OLS regression model on framing. When it came to the effect of people supporting the Black Lives Matter, the support showed a modest negative effect on the odds of the NYT quoting officials, which was consistent with original theoretical expectations. It also seemed to increase the degree of negativity through framing, which was a controversial finding. This, however, could be due to one specific complexity. The levels of support demonstrated the percentage of the US population supporting Black Lives Matter on a particular date. If there was a possibility that journalists would reflect on the levels of support much later, then this would explain such a controversial result.

The findings, despite the statistical imperfections, reveal several interesting trends and important challenges that future scholars with similar research goals need to overcome. The statistical significance of learning in hints at the importance of paying closer attention to the control variables included in the model, the effect of which might end up being more profound than that of independent variables. Unspecified confounders, like police brutality and individual biases, might also affect the results: not taking into consideration such important factors earlier might end up in a non-linear relationship or model misspecification. Finally, the conventional methods of testing linear relationship or probability, might not work well with the data of smaller sample size: for this reason, methods like negative binomial regression and zero-inflated models (see **Appendix A**) were not taken into consideration.

The findings, despite insufficient evidence on the presence of the protest paradigm in the NYT coverage of George Floyd protests in 2020, reveal several important theoretical implications. The role of time passed since the first day of the protests reveal how learning might affect the chances of the coverage to be negative more prominently than the expected predictors of violence and popularity. While such findings do not contradict the challenge to the protest paradigm based on ideological alignment (Weaver and Scacco, 2013), they are at odds with the arguments of Boyle et. al (2012) stressing the importance of tactics over the goals of the protest. Given the results of linear regression with clustered robust standard errors (Appendix C), the role of violence during protests, which directly relates to tactics, did not have a substantive and statistically significant effect on the response variables.

Regardless of the unexpected complications and mostly unconfirmed theoretical suggestions, I would still encourage the media and communication scholars to probe further into the reasons behind the tone of the coverage. It would be especially exciting to see a study of similar design but with a much bigger sample and unrestricted with time, resources, and level of programming mastery. Thankfully, the datasets organized by the Armed Conflict Location and

Event Data (ACLED) project and CIVIQS analytics company allow the inquiry with a profoundly larger time frame, now that the Black Lives Matter protests are ongoing to this day.

# Appendix A: Extra Models

In the late stages of the research, the distribution of response variable was revealed to be highly skewed to the right, while containing a lot of zero values. Before I decided to have two models for my analysis, I wanted to try out several other regression models that were reported to specialize when dealing with excess zeroes. Those were: Poisson regression, negative binomial regression, and zero-inflated models. Below, I list the reasons of why a particular model did not suit the analysis for my research:

- Poisson: according to UCLA, Poisson regression was not dealing well with the issue of excess zeroes, as opposed to happen-to-be zeroes. When tried out in RStudio, the values of null and residual deviance equaled to ~3000, which indicated a gross over-dispersion, as the number of degrees of freedom was substantively lower (~160). The common recommendation was to try out the negative binomial or zero-inflated models.
- Negative binomial: the codebook of UCLA Statistical Consulting strongly advised against using negative binomial models with smaller samples. Since my sample had a rather small number of observations, I decided to move on to the other alternatives.
- Zero-inflated models: similarly to the case of negative binomial models, the codebook advised against applying these type of models to smaller samples. Apart from that, when trying out the model in RStudio, the pseudo R-Squared for zero-inflated and hurdle models was equal to ~0.93, which seemed like an impossible estimate. Theoretically, it would have been challenging to justify the division of zeroes into happen-to-be and excess zeroes, as it did not suit the theory.

# Appendix B: Exponentiated Effects

This section contains the graphs for the effect of statistically significant predictors on the exponentiated values of negative frames in OLS regression model for framing (4.2.1). After exponentiation of log number of negative frames, all statistically significant predictors lose a linear relationship with the response variable for an unknown reason. In the case of the violence variable, which was log-transformed as well, the values were exponentiated in the same fashion. The graphs for exponentiated effect take place in the same order as the graphs for log effect.

Figure B.1. The exponentiated effect of support on predicted number of negative frames in the NYT coverage of the first two weeks of George Floyd protests in 2020.

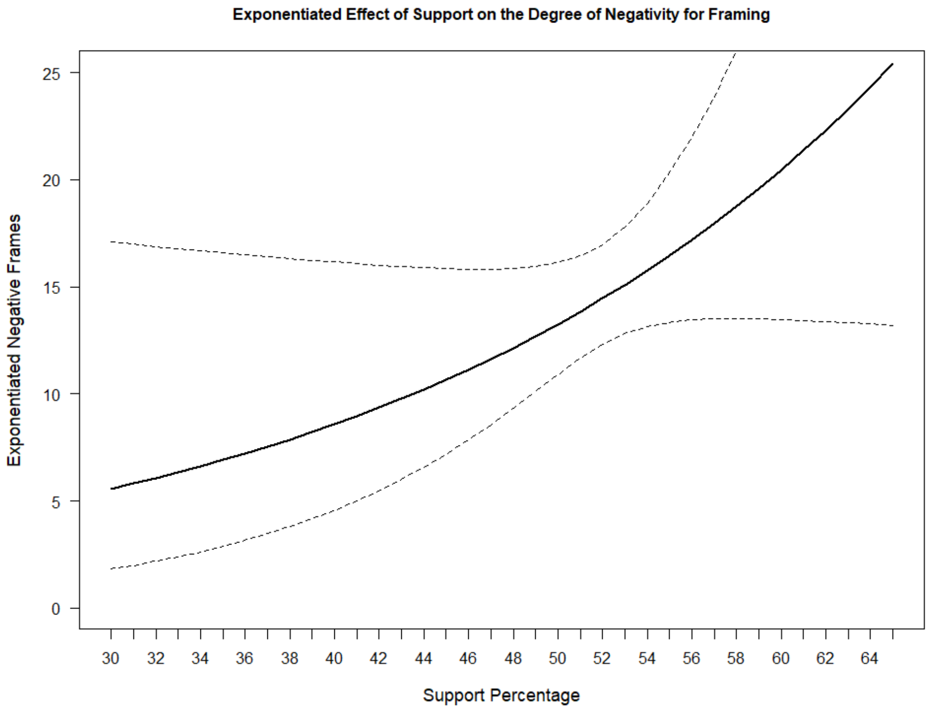


Figure B.2. The exponentiated effect of neutral attitudes on predicted number of negative frames in the NYT coverage of the first two weeks of George Floyd protests in 2020.

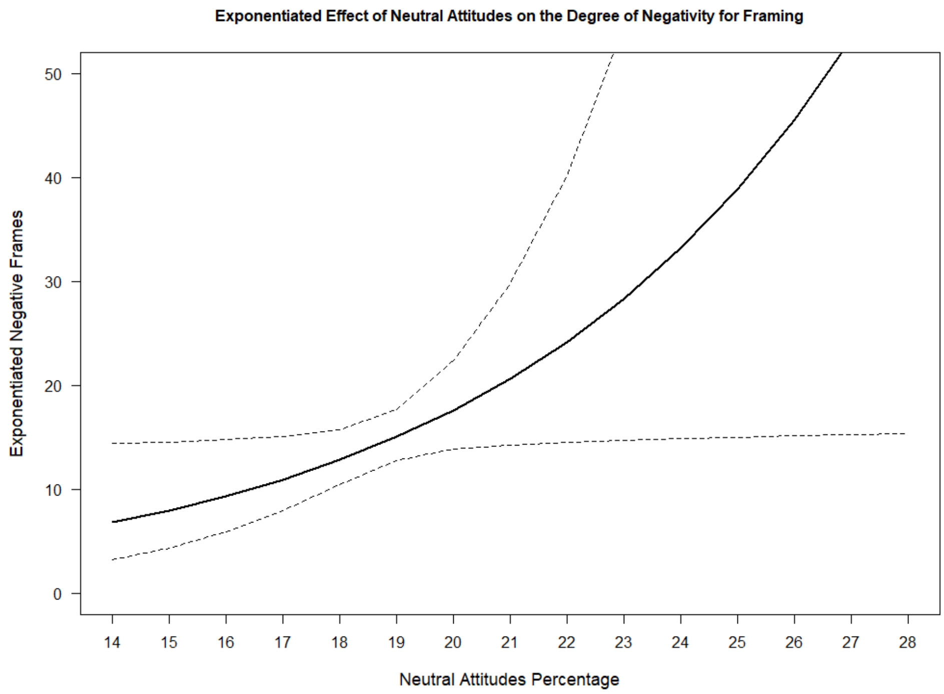



Figure B.3. The exponentiated effect of violence on predicted number of negative frames in the NYT coverage of the first two weeks of George Floyd protests in 2020.

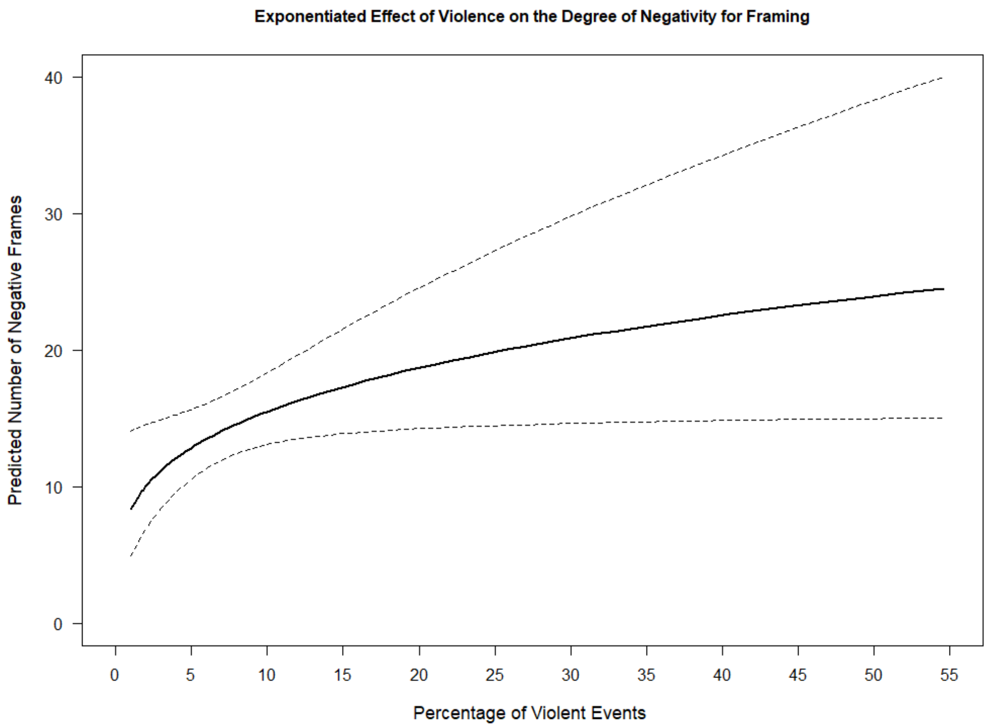
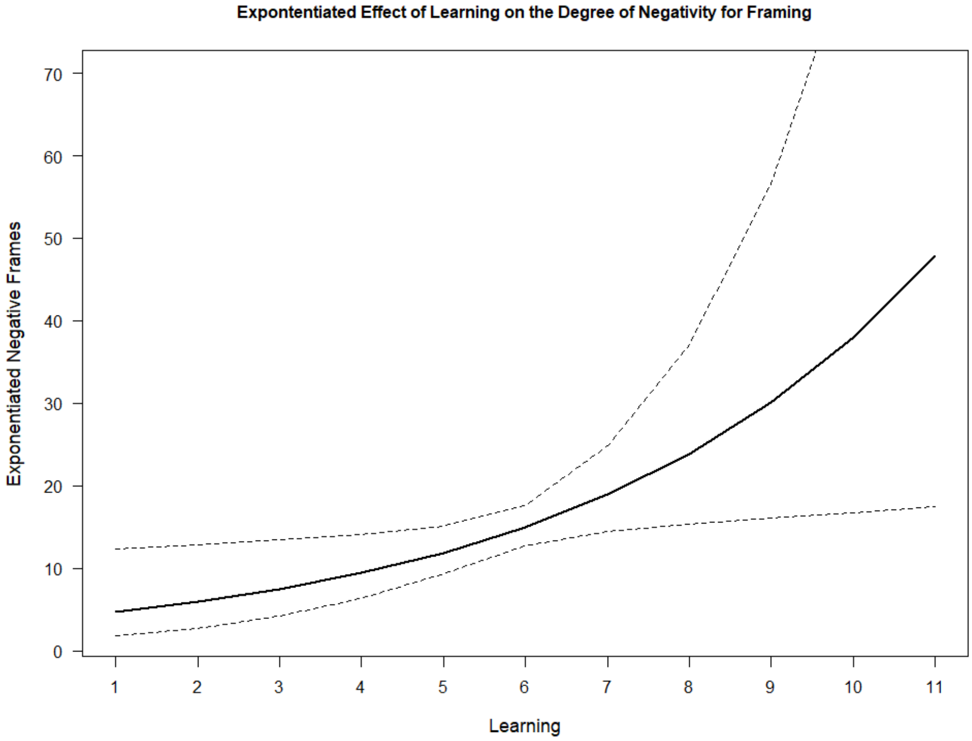


Figure B.4. The exponentiated effect of learning on predicted number of negative frames in the NYT coverage of the first two weeks of George Floyd protests in 2020.



# Appendix C: Clustered Standard Errors

As mentioned in the discussion chapter, due to peculiarities of data collection and organizing procedure, it was necessary to run clustered standard errors with each model as a robustness check. In order to avoid observations being mutually dependent on each other, I clustered each model by the state. The coefficients for each model after applying clustered standard errors can be found below in the relevant order. As can be seen in figures, while the coefficients experienced slight changes in all models, the only coefficient that was noticeably affected was the violence variable in OLS framing model (figure C3). After applying clustered errors, the positive effect of violence on the predicted log number of negative frames became of much less statistical significance (the p-value of 0.13), than earlier (the p-value of 0.028).

Figure C.1. The coefficients for the logit model on framing after applying clustered standard errors.

```
> summary(FramingLogitModel_c)
                  Estimate Std. Error    z value    Pr(>|z|)
(Intercept)    0.958551234 4.00249085  0.2394887 0.810726676
SupportCount   0.027309289 0.04571565  0.5973728 0.550258494
NeutralCount  -0.044701699 0.10395219 -0.4300217 0.667179846
EventsNo       0.001560957 0.01072018  0.1456092 0.884229878
Learning      -0.246812539 0.09043897 -2.7290507 0.006351693
ViolenceCount  0.016705947 0.01867048  0.8947786 0.370905456
```

Figure C.2. The coefficients for the logit model on sourcing after applying clustered standard errors.

```
> summary(OffLogitModel_c)
                  Estimate Std. Error    z value   Pr(>|z|)
(Intercept)    7.399392234 5.04725755  1.4660223 0.14264221
SupportCount  -0.073714943 0.04752841 -1.5509658 0.12090987
NeutralCount  -0.141905294 0.12324551 -1.1514033 0.24956634
EventsNo       0.022347772 0.01119147  1.9968582 0.04584059
Learning      -0.279949758 0.13491875 -2.0749507 0.03799111
ViolenceCount -0.002249488 0.02233843 -0.1007004 0.91978833
```

Figure C.3. The coefficients for the OLS regression model on framing after applying clustered standard errors.

```
> summary(FramingOLS_c)
R^2= 0.13126

                Estimate  Std. Error    t value    Pr(>|t|)
(Intercept)  -4.372643371 2.730819272 -1.6012203 0.10932812
SupportCount  0.043387503 0.025252890  1.7181203 0.08577467
NeutralCount  0.158190894 0.066388574  2.3828030 0.01718139
EventsNo     -0.005506322 0.007049847 -0.7810555 0.43476983
Learning      0.232086846 0.107652996  2.1558791 0.03109311
ViolenceLOG   0.269940733 0.180981308  1.4915393 0.13581996
```

Figure C.4. The coefficients for the OLS regression model on sourcing after applying clustered standard errors.

```
> summary(OfficialsOLS_c)
R^2= 0.10863

                 Estimate  Std. Error    t value  Pr(>|t|)
(Intercept)   4.495411646 3.378320805  1.3306645 0.1832994
SupportCount -0.026358314 0.024249523 -1.0869622 0.2770535
NeutralCount -0.055982447 0.089226104 -0.6274223 0.5303825
EventsNo     -0.001059505 0.005237539 -0.2022905 0.8396896
Learning      0.036096863 0.126355643  0.2856767 0.7751258
ViolenceLOG  -0.020969257 0.125488014 -0.1671017 0.8672901
```

# Reference List


Allison, Paul D. 2014. "Measures of Fit for Logistic Regression". https://statisticalhorizons.com/wp-content/uploads/GOFForLogisticRegression-Paper.pdf

BBC. 2020. "George Floyd Death: Ex-officer Charged with Murder in Minneapolis". https://www.bbc.com/news/world-us-canada-52854025

Benoit, Ken. 2011. "Linear Regression Models with Logarithmic Transformations". https://kenbenoit.net/assets/courses/ME104/logmodels2.pdf

Boyle, Michael P., and Cory L. Armstrong. 2009. "Measuring Level of Deviance: Considering the Distinct Influence of Goals and Tactics on News Treatment of Abortion Protests". https://doi.org/10.1080/15456870903156134

Boyle, Michael P., Douglas M. McLeod, and Cory L. Armstrong. 2012. "Adherence to the Protest Paradigm: The Influence of Protest Goals and Tactics on News Coverage in U.S. and International Newspapers". The International Journal of Press/Politics 17, no. 2: 127–44. https://doi.org/10.1177/1940161211433837

Boykoff, Jules. 2006. "Framing Dissent: Mass-Media Coverage of the Global Justice Movement". New Political Science, 28. https://doi.org/10.1080/07393140600679967

Bray, Mark. 2012. "Confronting the Many Men in Suits: Rethinking the 'Positive' Coverage of Occupy Wall Street." Critical Quarterly 54 (2): 5–9. https://doi.org/10.1111/j.1467-8705.2012.02049.x

CIVIQS. 2020. "Black Lives Matter. Registered Voters". https://civiqs.com/results/black_lives_matter

CNN. 2020. "Top US General says Iran Tried to Kill US Troops as Some Administration Officials Believe They Purposely Missed". https://edition.cnn.com/2020/01/08/politics/trump-iran-retaliation-missile-attacks/index.html

Dixon, Travis, and Cristina Azocar. 2006. "The Representation of Juvenile Offenders by Race on Los Angeles Area Television News." The Howard Journal of Communications 17: 143–161. https://doi.org/10.1080/10646170600656896

Dixon, Travis, and Daniel Linz. 2000a. "Overrepresentation and Underrepresentation of African Americans and Latinos as Lawbreakers on Television News." Journal of Communication 50 (2): 131–154. https://doi.org/10.1111/j.1460-2466.2000.tb02845.x

Dixon, Travis, and Daniel Linz. 2000b. "Race and the Misrepresentation of Victimization on Local Television News." Communication Research 27: 547–573. https://doi.org/10.1177%2F009365000027005001

Elmasry, Mohamad H. and Mohammed el-Nawawy. 2017. "Do Black Lives Matter?" Journalism Practice, 11:7, 857-875.

Financial Times. 2020. "Coronavirus: US death toll surpasses 83,000 — as it happened". https://www.ft.com/content/0187857a-3d2d-37e1-9749-d9789ac08591

FiveThirtyEight. 2020. How Americans Feel About George Floyd's Death And The Protests. https://fivethirtyeight.com/features/how-americans-feel-about-george-floyds-death-and-the-protests/

Gamson, William A., and Gadi Wolfsfeld. 1993. "Movements and Media as Interacting Systems." The Annals of the American Academy of Political and Social Science 528 (1993): 114-25. http://www.jstor.org/stable/1047795

Garrett, Kelly R., Brian E. Weeks, and Rachel L Neo. 2016. "Driving a Wedge Between Evidence and Beliefs: How Online Ideological News Exposure Promotes Political Misperceptions". Journal of Computer-Mediated Communication, 21: 331-348. https://doi.org/10.1111/jcc4.12164

Gerring, John. 2008. Case Selection for Case-Study Analysis: Qualitative and Quantitative Techniques. The Oxford Handbook of Political Methodology. https://doi.org/10.1093/oxfordhb/9780199286546.003.0028
Gitlin, Todd. 1980. "The Whole World is Watching: Mass Media in the Making & Unmaking of the New Left". Berkeley: University of California Press.
Gottlieb, Julian. 2015. "Protest News Framing Cycle: How The New York Times Covered Occupy Wall Street". International Journal of Communication, 9. https://ijoc.org/index.php/ijoc/article/view/2880
Grimmer, Justin, and Brandon M. Stewart. 2017."Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts" in Political Analysis, 21(3). https://doi.org/10.1093/pan/mps028
Hopkins, Daniel J., Eunji Kim, and Soojong Kim. 2017. "Does Newspaper Coverage Influence or Reflect Public Perceptions of the Economy?" Research & Politics.
Kilgo, Danielle K., and Summer Harlow. 2019. "Protests, Media Coverage, and a Hierarchy of Social Struggle". The International Journal of Press/Politics 24, no. 4: 508–30. https://doi.org/10.1177/1940161219853517
Kilgo, Danielle K., and Rachel R. Mourao. 2019. "Media Effects and Marginalized Ideas: Relationships Among Media Consumption and Support for Black Lives Matter". International Journal of Communication 13: 19.
Kilgo, Danielle K., and Rachel R. Mourão. 2021. "Protest Coverage Matters: How Media Framing and Visual Communication Affects Support for Black Civil Rights Protests". Mass Communication and Society. https://doi.org/10.1080/15205436.2021.1884724
Kim Kisun, and Saif Shahin. 2020. "Ideological Parallelism: Toward a Transnational Understanding of the Protest Paradigm". The Social Movement Studies, 19:4, 391-407.
Kishi, Roudabeh and Sam Jones. 2020. "Armed Conflict Location and Event Data Project. Demonstrations and Political Violence in America: New Data for Summer 2020". https://acleddata.com/acleddatanew/wpcontent/uploads/2020/09/ACLED_USDataReview_Sum2020_SeptWebPDF_HiRes.pdf
LA Times. 2020. "George Floyd protests have created a multicultural movement that's making history". https://www.latimes.com/california/story/2020-06-07/george-floyd-protests-unite-black-activists-new-allies
McLeod, Douglas M., and James K. Hertog. 1999. "Social Control and the Mass Media's Role in the Regulation of Protest Groups: The Communicative Acts perspective". Mass Media, Social Control and Social Change, 305-30.
McLeod, Douglas M., and Detenber, Benjamin. 2006. "Framing Effects of Television News Coverage of Social Protest". Journal of Communication. Vol. 49, no. 3.
McLeod, Douglas M. 2007. "News Coverage and Social Protest: How the Media's Protest Paradigm Exacerbates Social Conflict". Journal of Dispute Resolution, Vol. 2007, Issue 1.
Meyer, David S. 2003. "How Social Movements Matter". Contexts 2, no. 4: 30–35. https://doi.org/10.1525/ctx.2003.2.4.30
Molotch, Harvey. 1979. "Media and Movements in The Dynamics of Social Movements". ed. Mayer N. Zald and John McCarthy, Cambridge, MA: Winthrop.
NBC News. 2020. "First Came a Pandemic. Then, Looting. Small Businesses Pick up the Pieces as Their Debt Mounts". https://www.nbcnews.com/business/business-news/first-came-pandemic-then-looting-small-businesses-pick-pieces-their-n1224776
NPR. 2020. "Protests Over George Floyd's Death Escalate into Violence, Destruction". https://www.npr.org/sections/pictureshow/2020/05/28/864559449/photos-protests-over-george-floyds-death-grow-violent-it-is-how-we-express-pain?t=1615138540931
Pew Research Center. 2014. "Political Polarization & Media Habits". https://www.journalism.org/2014/10/21/political-polarization-media-habits/

Pew Research Center. 2016. "How Americans View the Black Lives Matter Movement". https://www.pewresearch.org/fact-tank/2016/07/08/how-americans-view-the-black-lives-matter-movement/
Politico. 2020. "Trump Acquitted on Impeachment Charges, Ending Gravest Threat to His Presidency". https://www.politico.com/news/2020/02/05/trump-impeachment-vote-110805
Reuters Institute. 2020. "Reuters Institute Digital News Report 2020". https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2020-06/DNR_2020_FINAL.pdf
Scheufele, Dietram A., and David Tewksbury. 2006. "Framing, Agenda Setting, and Priming: The Evolution of Three Media Effects Models". Journal of Communication, 57,1. https://doi.org/10.1111/j.0021-9916.2007.00326.x
Schoonvelde, Martijn. 2020. "Automated Text Analysis in Political Science: Preprocessing Data, Going from Text to Numbers". School of Politics and International Relations, UCD.
Strömbäck Jesper. 2012. "The Media and Their Use of Opinion Polls: Reflecting and Shaping Public Opinion". Opinion Polls and the Media, 1-22. Palgrave Macmillan.
Tarrow, Sidney G. 1998. "Power in Movement: Social Movements and Contentious Politics". Cambridge: Cambridge University Press.
Tenenboim-Weinblatt, Keren. 2014. "Producing Protest News: An Inquiry into Journalists' Narratives". The International Journal of Press/Politics 19, 410–29. https://doi.org/10.1177/1940161214540941
The Guardian. 2020. "George Floyd's Daughter and Her Mother Make First Public Appearance."https://www.theguardian.com/us-news/2020/jun/02/george-floyd-protests-curfews-eighth-day
The Guardian. 2020. "Hundreds Demand Justice in Minneapolis after Police Killing of George Floyd". https://www.theguardian.com/us-news/2020/may/26/george-floyd-killing-minneapolis-protest-police
The New York Times. 2020. "Protests Continue to Rage After Death of George Floyd". https://www.nytimes.com/2020/05/28/us/george-floyd-national-guard.html
The New York Times Company. 2021. "The New York Times Company Reports 2020 Fourth-Quarter and Full-Year Results and Announces Dividend Increase". https://nytco-assets.nytimes.com/2021/02/Press-Release-12.27.2020-Final-for-posting.pdf
Tillery, Alvin B. Jr. 2019. "What Kind of Movement is Black Lives Matter? The View from Twitter". Journal of Race, Ethnicity and Politics, 2, 297-323. https://doi.org/10.1017/rep.2019.17
Tjur, Tue. 2009. "Coefficients of Determination in Logistic Regression Models—A New Proposal: The Coefficient of Discrimination". The American Statistician, 63, (4). http://www.jstor.org/stable/25652317
UCLA Institute for Digital Research and Education. "FAQ: How Do I Interpret a Regression Model when Some Variables are Log Transformed?". https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faqhow-do-i-interpret-a-regression-model-when-some-variables-are-log-transformed/
Vliegenthart, Rens, Stefaan Walgrave, Ruud Wolters and Swen Hutter. "The Media as a Dual Mediator of the Political Agenda–Setting Effect of Protest. A Longitudinal Study in Six Western European Countries". Social Forces, 95(2).
Washington Post. 2021. "Some Police Deaths are More Worrisome for Fox News Than the Others". https://www.washingtonpost.com/politics/2021/02/03/some-police-deaths-are-more-worrisome-fox-news-than-others/
Watson, Amy. 2021. "Weekday Circulation of The New York Times from 2000 to 2020". Statista. https://www.statista.com/statistics/273503/average-paid-weekday-circulation-of-the-new-york-times/

Weaver, David A., and Joshua M. Scacco. 2013. "Revisiting the Protest Paradigm: The Tea Party as Filtered through Prime-Time Cable News". International Journal of Press/Politics 18: 61–84. https://doi.org/10.1177/1940161212462872

Wright, Wynne, and Taylor Reid. 2011. "Green Dreams or Pipe Dreams? Media Framing of the U.S. Biofuels Movement." Biomass & Bioenergy 35 (4): 1390–99. https://doi.org/10.1016/j.biombioe.2010.07.020