

Description Bias in 2020 Black Lives Matter Uprisings’ News Coverage

Kai Nham

University of California, Berkeley School of Information

knham@berkeley.edu

Abstract

This paper applies several natural language processing techniques in order to analyze the media coverage of the 2020 Black Lives Matter uprisings. This is done through formalizing “description bias” in news coverage of protests. It was found that police were afforded more agency but less power through the coverage, which continues to bias news coverage towards the police, since protesters were written as powerful forces that needed to be controlled and contained. This justifies the use of militarized force and continues to reinforce public approval for state repression of Black voices.

1 Introduction

Nina Pop. Breonna Taylor. Tony McDade. George Floyd. Ahmaud Arbery. These are the people, among many others, whose deaths at the hands of white supremacy and state violence roused mass uprisings across the United States and around the world in 2020.

While one of the goals of mass demonstrations can be to attract media attention, it is part of the broader political strategy to place acute pressure on decision makers to act in alignment with a political

agenda set forth by community organizers. However, both historically and contemporarily, the media has played a major role in framing and shaping popular discourse around the agendas set forth by community organizers and activists. As Smith et al. (2001) describes, the media controls the narrative through two major biases: selection bias and description bias. Selection bias narrows the pool of political uprisings and demonstrations that are deemed “newsworthy” by editorial staff. But even if a demonstration is selected for coverage, it is subject to description bias, where “protest activities or organizers are often portrayed in a manner that reporters believe will appeal to a mass audience” (Smith et al, 2001).

To formalize description bias, I utilize several natural language processing techniques to provide an analysis of description bias in mainstream media coverage. First, I look at the most common words in the corpus of news articles covering the 2020 Black Lives Matter uprisings. Second, I look at the cosine similarity between specific words of interest, which were pulled from the most common words in the corpus. Lastly, and perhaps most interestingly, I apply connotation frames of agency and power from Sap et al. (2017)’s lexicon to the major

actors of the uprisings. In the context of the 2020 Black Lives Matter uprisings, I choose to focus on two major actors: the police and the protesters. The police play an interesting role in these protests because they not only serve as the oppositional reaction to control and contain the protests but also are the institution being directly called for accountability in the deaths of George Floyd, Breonna Taylor, and others. Protesters are important actors because they collectively represent and form the social movement as a

In summary, this quantitative analysis provides insight into the ways that mainstream news outlets portrayed the 2020 Black Lives Matter uprisings, including the description bias towards police power and containment over the exercising of agency of Black communities.

2 Background

By choosing a particular verb, writers connote rich information about the entities that interact through the verb (Rashkin et al., 2016). Rashkin et al. (2016) developed connotation frames, which created a representation framework to organize these depths and dimensions, moving past the denotational meanings of language. Connotation frames provide a way to do a deep analysis of the many layers of value-laden interpretation embedded in writing choices.

Building upon this work, Sap et al (2017) developed connotation frames of power and agency, which serve as a way “to model how different levels of power and agency are implicitly projected on actors through their actions” (Sap et al., 2017). These connotation frames were developed through annotation by three AMT crowdworkers with placeholders to

control for gender bias (i.e. “*X* obeyed *Y*” as a sample task). The primary focus of the paper was specifically to identify gender bias in movie scripts through observing different levels of agency and power given to nominal subjects versus themes through predicate choice. Power is attributed to an agent when verbs are used that imply a level of control over the theme, while agency is attributed to an agent when verbs are used to imply an agent is “powerful, decisive, and capable of pushing forward their own storyline” (Sap et al., 2017).

The development of these connotation frames allow for further exploration and analysis of other written work to expose potential biases. As such, the application of Sap et al. (2017)’s frames to the 2020 Black Lives Matter uprisings’ news coverage can provide insight into description bias that could potentially hinder or placate agenda setting and building for progressive social movements.

3 Methods

Foremost, pulling together a sizable corpus was required for analysis of the 2020 Black Lives Matter uprisings. This was done by scraping search results for “George Floyd protests”¹ on mainstream news sites, either directly through the site or using the NewsAPI. The final corpus consists of 590 news articles up until June 28, 2020 from mainstream news sources including Associated Press, CNN, San Francisco Chronicle, Washington Post, and USA Today.

¹ I want to acknowledge and problematize that George Floyd is the one often attached to these uprisings, when it has been the tremendous and constant loss of Black life – and a commitment to ensure not another would be taken – that moves these uprisings.

In approaching this analysis, SpaCy’s English medium core model was used to conduct tokenization, dependency parsing, and cosine similarity analysis. The SpaCy model is a multi-task convolutional neural network trained on OntoNotes 5 – a large annotated corpus composed of various genres of text from news to broadcast conversations – and GloVe vectors trained on Common Crawl. This model was chosen because it was trained on a large news corpus, and thus could generalize well to the relatively small dataset on the 2020 Black Lives Matter uprisings. Additionally, the capability of the model to perform multiple natural language processing tasks with relatively high accuracy allowed for rich and nuanced analysis on the news coverage.

Once the dependency parsing was complete for all of the articles, verbs with nominal subjects that matched our actors of interest (e.g. “police” and “protesters”) were pulled and lemmatized in order to apply connotation frames of power and agency. There are a total of 2155 verbs, originally labeled with strings to indicate positive, negative, or equal agency and/or power where applicable. These labels were converted into 1, -1, 0, or NaN respectively in order to make quantifiable distinctions between the two actors’ connotative power and agency.

4 Results & Discussion

First, a simple word count, excluding stop words and punctuation, was conducted. The word “police” was mentioned 10419 times in the 590 articles and was the second most frequent word in the corpus, while “protesters” only appeared 3627 times and was the eighth

most frequent word. Nearly three times the amount of mentions, police were undoubtedly a central pillar of the coverage in the 2020 Black Lives Matter uprisings; however, since the demands of the protesters centered around defunding the police and other abolitionist reforms, it is unclear whether from word count alone whether they were nominal subjects at that high of a frequency.

Second, similarities between words of interest were calculated. Interestingly, out of “protests,” “protesters,” and “police,” “violence” and “violent” were most similar to “protests.” Indicative of its similarity to the adjective form, protests have been characterized as violent, despite individual actors (namely, protesters) having greater cosine dissimilarity. Although individual actors are not portrayed as violent, this indicates a potential tendency to see protests as mass, spontaneous disruptions that enact violence en masse randomly rather than as coordinated and strategic actions for agenda building.

Moreover, the great dissimilarity between “police” and “peaceful” with the simultaneous ambivalent similarity ranking between “police” and “violent” continues to paint police as neutral arbiters of law and order. While they are not definitively not peaceful, they are not readily described as violent, although more so than protesters. This portrayal is furthered by the analysis of connotation frames.

Word 1	Word 2	Similarity
Protests	Violence	0.54861
Police	Violence	0.54212
Protests	Violent	0.52590
Police	Violent	0.48235
Protesters	Violence	0.47344

Protesters	Violent	0.46466
Protests	Peaceful	0.44063
Protesters	Peaceful	0.42890
Police	Peaceful	0.27409

Table 1: The cosine similarities between words of interest based on some of the words found in the top 50 most frequent words in the corpus.

After dependency parsing, it was found that protesters appeared as nominal subjects in 1245 separate occurrences in the corpus, whereas police appeared 1528 times. This is important to note, since protesters are not even written as agents as frequently as police were in the articles, despite the fact that the articles were about the protests themselves. Since police represent the state, they are prioritized in the news gathering process, since most media relies on “official” or government sources due to their convenience, as well as because “official sources limit risks of alienating influential elites” (Smith et al., 2001). The media, after all, serves as a dual role in policing – one that fosters the notion of policing as essential by reproducing public consent and simultaneously portrays policing as threatening through reporting instances of wrongdoing (Higgins, 2020). Coverage, however, has a tendency to favor the latter through framing of the individual culpability of a “few bad apples,” rather than a manifestation of structural police power (Higgins, 2020). This framing of individual responsibility through portrayal of the problem without reference to the systemic causes or the dramatization of the protest, detracts from the substance of protester critiques and thus, discourages critical engagement of audiences (Smith et al., 2001).

Moreover, the collective feminist and non-hierarchical nature of the 2020 Black Lives

Matter uprisings produced no obvious traditional “leaders [who] provide the most appealing media images and rhetoric... to gain media coverage” (Smith et al., 2001). Additionally, for valid fears of lashback, most protesters are advised not to show their face or talk to the media. This leads to a lack of traditionally deemed “credible” sources that can speak individually to the power and demands of the movements, leading to less active framing of protesters.

This analysis is furthered by the application of connotation frames of power and agency. After performing a one-tailed t-test analysis to compare police agency versus protester agency and police power versus protester power, it was found that police were written with more agency, while protesters were written with more power. Both differences were found to be statistically significant.

Null Hypothesis	Statistic	p-value
pol_agency <= pro_agency	2.57261	0.00507
pol_power >= pro_agency	-2.11423	0.01730

Table 2: The results of the one-tailed t-tests conducted on police agency versus protester agency and police power versus protester power. A standard alpha of 0.05 was used to determine statistical significance.

The question arises: how police are afforded more agency in their actions, yet they are portrayed as not as powerful despite being fully armed and supported by state power? This dynamic reveals a description bias that favors and safeguards the institution of policing, despite the repression and violence enacted in the deaths of George Floyd and others, as well as in response to the protests. Because the

characterization of the protests is more aligned with “violence,” the clashes between police and protesters are dramatized and often afford agency to police to contain and control but still portray protesters as powerful actors that the police are struggling against. Additionally, the portrayal of protesters as powerful but not as full of agency as police, continues to reify ideas of the spontaneous and disruptive nature of these mobilizations, rather than concerted and strategic efforts that stand in a long legacy of Black struggle against state violence and repression.

5 Conclusion

Through the application of several analytical techniques harnessing natural language processing, a quantitative analysis of 2020 Black Lives Matter uprisings’ mainstream news coverage was developed. Specifically, through looking at word counts, cosine similarity between words, and connotation frames of power and agency, it was found that there is a description bias towards police in coverage.

References

Kat Higgins. 2020. Why police violence is a media problem. June 9, 2020.

Hannah Rashkin, Sameer Singh, and Yejin Choi. 2016. Connotation Frames: A Data-Driven Investigation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 311–321, Berlin, Germany, August. Association for Computational Linguistics.

Maarten Sap, Marcella Cindy Prasettio, Ari Holtzman, Hannah Rashkin, and Yejin Choi. Connotation Frames of Power and Agency in Modern Films. :6.

Jackie Smith, John D. McCarthy, Clark McPhail, and Boguslaw Augustyn. 2001. From Protest to Agenda Building: Description Bias in Media Coverage of Protest Events in Washington, D.C. *Social Forces*, 79(4): 1397–1423.