Kai McNamee
Gov 51
May 9, 2022
GitHub repository

The Effect of News Coverage on Police Favorability

How does news media affect the public's perception of the police? This analysis aims to answer this question by exploring how the text of news articles influences readers' attitudes towards the police. As police reform and abolition have become increasingly mainstream topics in American politics, the media plays an important role in informing and influencing the public. This paper is inspired by Reny and Newman's (2021) study on the impact of social protest against police violence, which used regression discontinuity design to illustrate how the beginning of Black Lives Matter protests in the summer of 2020 was correlated with a sharp decline in favorability towards the police. I hope to expand on their analysis and further dissect how media consumption influences attitudes towards the police by applying sentiment analysis to New York Times articles. I hypothesized that individuals who consume news with more positive sentiment scores will report more favorable attitudes towards the police. The results support my hypothesis, showing a slight positive relationship between New York Times article sentiment scores and favorable attitudes towards the police, but the relationship is likely not causal.

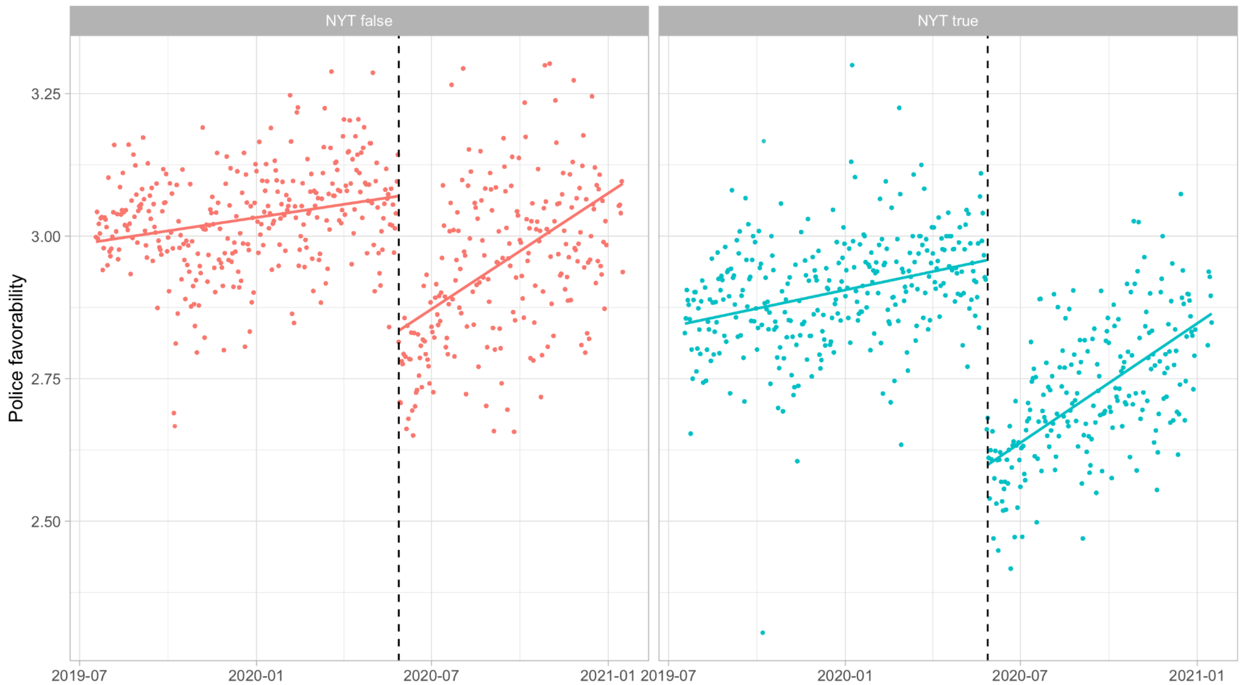
Data and Methods

Dependent variable: Attitudes towards the police.

The dependent variable of interest, individuals' attitudes towards the police, comes from the Nationscape survey, a cross-sectional opinion survey conducted between July 18, 2019, and November 2, 2021, by UCLA and the Democracy Fund. With around 6,000 responses every week, the dataset is a representative sample of US residents by gender, census region, race, income, education, age, and more (Tausanovitch et al 2019). To gauge individuals' attitudes towards the police, respondents were asked how favorable their impressions are of the police. Responses are stored in the variable "police," and coded as (4) "very favorable," (3) "somewhat favorable," (2) "somewhat unfavorable," or (1) "very unfavorable." (Note: the original data codes favorability towards the police such that 1 is most favorable, but I've inverted the scale for ease of interpretation). Respondents were also aksed where they might have "head news about politics in the past week" (New York Times, Fox, CNN, Facebook, etc.) — the "nyt" variable in analysis is coded as TRUE if the respondent consumed political news in the last week from the Times, or FALSE if not.

Figure 1 illustrates the average daily police favorability over time for individuals, with a discontinuity at August 12, 2020, the same day Reny and Newman's study used to indicate the beginning of Black Lives Matter protests. The data is split into respondents who indicated they had read political news from the Times in the last week ("NYT true") and those who did not ("NYT false").

FIGURE 1: Average daily police favorability



Independent variable: New York Times article sentiments

The independent variable of interest is the sentiment score of New York Times articles. To generate sentiment scores for Times articles, I first used the Times' Article Search API to retrieve all articles returned under a query for "police" in the relevant sections during the span of the Nationscape data. Article headlines and lead paragraphs were then assigned sentiment scores using the R package sentimentr's lexicon-based sentiment analysis (Rinker, 2018). The approach involves using an expert-coded dictionary of sentiment scores (positive scores for positively coded words, negative scores for negatively coded words) to generate an average sentiment score for a body of text.

Since the Times Article Search API does not return the full text of articles, I approximated the sentiment of a given article with the average sentiment of its lead and headline, weighted by the number of words in each. To increase the chances of an article being related to readers' opinions towards the police, I filtered the search results to only include those that explicitly mention "police" in the headline or lead. This approach to simplifying the independent variable is justified on two main assumptions: (1) news articles include the most important information within the headline or lead paragraph (in line with the "inverted pyramid" style of news writing), so including more text introduces potentially irrelevant noise into the analysis; (2) the headline and lead paragraph are the most visible part of any news article shared online, meaning Nationscape respondents who indicated having consumed political news from the Times would have seen at minimum the text included in the analysis. This methodology is consistent with previous lexicon-based approaches to analyzing news media (Burscher et al, 2016).

To join the data, I averaged sentiment scores for all Times articles published in the week leading up to an individual's response on the Nationscape survey. This lagged sentiment mean is stored as "article_week" in my data.

I also included other control variables like party (Democrat, Republican, or Independent), race (White, Black, Asian, Native American, and other), and whether or not the respondent consumed political news from CNN or Fox in the prior week.

Results

The results of the analysis show a slight positive correlation between lagged sentiment scores and individuals' attitudes towards the police, an effect that is more pronounced for readers of the Times. Figure 2 illustrates the relationship between police favorability and lagged article sentiment.

FIGURE 2: Police favorability over article sentiment

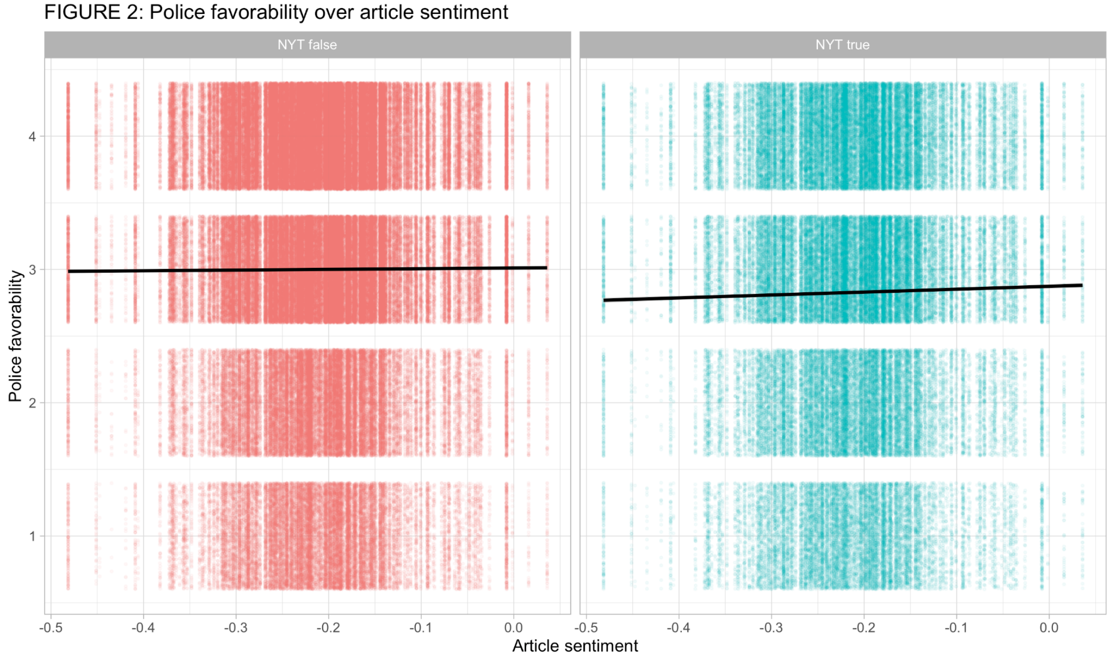


Figure 3 illustrates the average daily police favorability against lagged article sentiments, broken down by party.

FGURE 3: Average article sentiment and police favorability by day

Before and after the beginning of summer 2020 protests

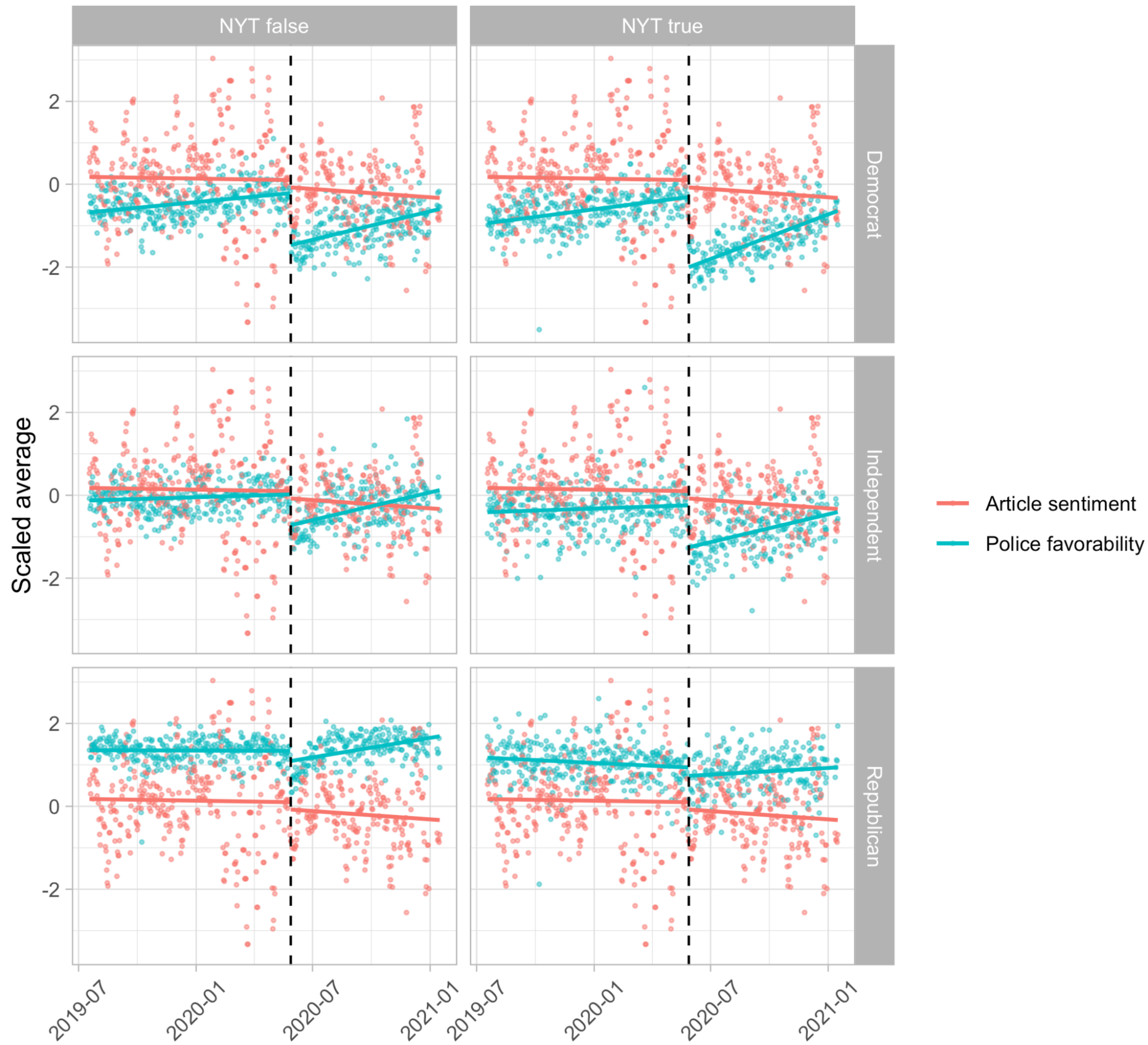


Table 1 illustrates an initial regression analysis, using the model specification for data subsetted to Times readers and Times non-readers:

$$Police\ favorability_i = \beta_0 + \beta_1 Article\ sentiment_i + \beta_2 Party_i + \beta_3 Gender_i + \beta_4 CNN_i + \beta_5 Fox_i + \varepsilon$$

Where police favorability indicates an individual i's favoarability towards the police and article sentiment indicates lagged article sentiment mean.

TABLE 1: Police favorability by reader status

	Police favorability	
	NYT readers	NYT non-readers
	(1)	(2)
Article sentiment	0.197^{***}	0.072^{***}
	(0.034)	(0.022)
Party: Independent	0.075^{***}	0.068^{***}
	(0.007)	(0.005)
Party: Republican	0.468^{***}	0.461^{***}
	(0.007)	(0.005)
Race: Black	-0.523^{***}	-0.670^{***}
	(0.008)	(0.006)
Race: Asian	-0.189^{***}	-0.249^{***}
	(0.011)	(0.009)
Race: Native American	-0.290^{***}	-0.340^{***}
	(0.023)	(0.015)
Race: Other	-0.310^{***}	-0.375^{***}
	(0.011)	(0.008)
Gender: Male	-0.025^{***}	-0.106^{***}
	(0.006)	(0.004)
CNN	-0.063^{***}	-0.110^{***}
	(0.006)	(0.004)
Fox	0.211^{***}	0.167^{***}
	(0.006)	(0.004)
Constant	2.792^{***}	2.981^{***}
	(0.010)	(0.006)
Observations	127,907	273,666
R^2	0.099	0.141
Adjusted R^2	0.099	0.141
Residual Std. Error	0.986 (df = 127896)	0.934 (df = 273655)
F Statistic	$1{,}406.586^{***}$ (df = 10; 127896)	$4{,}495.443^{***}$ (df = 10; 273655)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 2 illustrates the results of a similar model specification, but subsetted by party instead of reade status:

$$Police\ favorability_i = \beta_0 + \beta_1 NYT\ reader_i + \beta_2 Article\ sentiment_i + \beta_3 Race_i + \beta_6 Gender_i + \beta_7 CNN_i + \beta_8 Fox_i + \beta_9 NYT\ reader_i * \ Article\ sentiment_i + \ \varepsilon$$

Where NYT Reader indicates individual i's reader status (TRUE for reader, FALSE for non-reader), and NYT Reader * Article sentiment indicates the interaction between reader status and lagged sentiment mean.

TABLE 2: Police favorability by party

	Police favorability		
	Democrats	Republicans	Independents
	(1)	(2)	(3)
NYT reader	-0.078^{***}	-0.049^{***}	-0.074^{***}
	(0.014)	(0.015)	(0.018)
Article sentiment	0.110^{***}	0.031	0.077^{*}
	(0.039)	(0.034)	(0.043)
Race: Black	-0.561^{***}	-0.605^{***}	-0.753^{***}
	(0.006)	(0.014)	(0.010)
Race: Asian	-0.248^{***}	-0.231^{***}	-0.187^{***}
	(0.011)	(0.014)	(0.013)
Race: Native American	-0.351^{***}	-0.216^{***}	-0.384^{***}
	(0.023)	(0.022)	(0.022)
Race: Other	-0.345^{***}	-0.292^{***}	-0.386^{***}
	(0.010)	(0.014)	(0.011)
Gender: Male	-0.050^{***}	-0.116^{***}	-0.076^{***}
	(0.005)	(0.005)	(0.006)
CNN	-0.009	-0.179^{***}	-0.133^{***}
	(0.006)	(0.006)	(0.006)
Fox	0.165^{***}	0.171^{***}	0.222^{***}
	(0.006)	(0.005)	(0.006)
NYT reader*Article sentiment	0.181^{***}	0.057	0.064
	(0.063)	(0.067)	(0.077)
Constant	2.893^{***}	3.446^{***}	3.032^{***}
	(0.010)	(0.008)	(0.010)
Observations	154,361	131,895	115,317
R^2	0.054	0.044	0.072
Adjusted R^2	0.054	0.044	0.072
Residual Std. Error	0.990 (df = 154350)	0.868 (df = 131884)	0.987 (df = 115306)
F Statistic	889.656^{***} (df = 10; 154350)	609.502^{***} (df = 10; 131884)	897.911^{***} (df = 10; 115306)
Note:			$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 3 utilizes regression discontinuity to examine the relationship between police favorability and article sentiments before and after treatment date August 12, 2020. This regression draws from Reny and Newman's discontinuity design, with the addition of the Times article sentiment variable. I subsetted the data by party and used the model specification:

$$Police\ favorability_i = \beta_0 + \beta_1 Day_i + \beta_2 Post_i + \beta_3 Article\ sentiment_i + \beta_4 Day_i * Post_i + \beta_5 Article\ sentiment_i * Post_i + \varepsilon$$

Where Day indicates the date of individual i's response minus August 12, 2020 and Post indicates whether i's response was before (FALSE) or after (TRUE) August 12, 2020.

TABLE 3: NYT readers' police favorability by party (RDD)

	Police favorability		
	Democrats	Republicans	Independents
	(1)	(2)	(3)
Day	0.001***	-0.0002**	0.0001*
	(0.0001)	(0.0001)	(0.0001)
Post	-0.542***	-0.041	-0.271***
	(0.029)	(0.036)	(0.039)
Article sentiment	0.144**	0.007	-0.103
	(0.061)	(0.076)	(0.080)
Day*Post	0.001***	0.001***	0.001***
	(0.0001)	(0.0001)	(0.0002)
Article sentiment*Post	0.032	0.151	0.305**
	(0.113)	(0.140)	(0.153)
Constant	2.870***	3.262***	2.846***
	(0.017)	(0.020)	(0.022)
Observations	59,640	33,859	34,408
R^2	0.021	0.002	0.010
Adjusted R^2	0.021	0.001	0.010
Residual Std. Error	1.013 (df = 59634)	0.949 (df = 33853)	1.026 (df = 34402)
F Statistic	252.499*** (df = 5; 59634)	10.666*** (df = 5; 33853)	67.363*** (df = 5; 34402)
Note:			*p<0.1; **p<0.05; ***p<0.01

Conclusion

The results show a statistically significant relationship between lagged Times article sentiments and favorability towards the police in some model specifications, supporting my initial hypothesis. But due to conflicting results and the limitations of this analysis, I conclude the relationship is likely not causal. The article sentiment coefficient of Table 1 indicates that a change in lagged sentiment results in a statistically significant 0.197 change in police favorability, meaning that as article sentiment becomes more positive, attitudes towards the police become more positive. Similarly, Table 2 indicates a statistically significant interaction between Times reader status and lagged article sentiment — for Democrat Times readers, article sentiment has a 0.181 effect on police favorablity. These findings, however, are complicated by the fact that article sentiment has a statistically significant positive relationship with police favorability even among non-readers. Article sentiment coeeficients indicate significant positive correlations within all non-readers ($\beta = 0.072$, Table 1). Table 3 shows that article sentiment is again positively corelated with police favorability for Democrats, but the the effect of article sentiment after the treatment date (Article sentiment * Post) is not statistically significant. If the effect of lagged article sentiment were causal, I would not expect these relationships. The biggest limitation of this analysis that could be improved is the coding of the independent variable — the results suggest that the lexicon-based sentiment analysis is just reflecting the state of current events, rather than indicating a favorable or unfavorable sentiment towards the police as an entity. This would explain why some readers of the Times experience a greater decline in police favorability after the treatment date (Figure 3) — more informed survey respondents likely would have experienced a greater change in police favorability after the murder of George Floyd regardless of lagged sentiments. This would also explain the effect of lagged sentiment on non-readers' favorability towards the police, who might be influenced by other news sources that parallel Times sentimens due to shared subject matter.

Works Cited

Burscher, B., Vliegenthart, R., & Vreese, C. (2016). Frames beyond words: Applying cluster and sentiment analysis to news coverage of the nuclear power issue. *Social Science Computer Review*.

Reny, T., & Newman, B. (2021). The opinion-mobilizing effect of social protest against police violence: Evidence from the 2020 George Floyd protests. *American Political Science Review*. https://doi.org/10.1017/S0003055421000460

Rinker, T. (2018). Sentimentr. https://gist.github.com/democracyobserver/4aaee327a84c145cfbcf9292dbb09e3a

Tausanovitch, C., & Vavreck, L. (2020). Democracy Fund + UCLA Nationscape. https://www.voterstudygroup.org/data/nationscape