The Press and Peace, Examining Iraq War Coverage in Newspapers using BERT LLMs

by

Jakobe Bussey[1]

Professor Peter Arcidiacono, Faculty Advisor

Duke University

jzb4@duke.edu

April, 2024

Abstract:

This study utilizes state-of-the-art BERT (Bidirectional Encoder Representations from Transformers) models to perform sentiment analysis on *Wall Street Journal* and *New York Times* articles about the Iraq War published between 2002 and 2012 and further categorize them using advanced unsupervised machine learning techniques. By utilizing statistical analysis and quartic regression models, this paper concludes that the two newspapers report on the Iraq War differently, with both exhibiting a predominantly negative-neutral tone overall. Additionally, the analysis reveals significant fluctuations in negativity from both outlets over time as the war progresses. Furthermore, this study examines the objectivity of reporting between editorial and non-editorial articles, finding that non-editorials tend to report more objectively, and the neutrality of editorials remains relatively constant while the objectivity of non-editorials fluctuates in response to war events. Finally, the paper investigates variations in sentiment across different topics, uncovering substantial variations in positive, neutral, and negative sentiments across topics and their evolution over time.

Keywords: Machine Learning, Media, Natural Language Processing (NLP)

JEL Classification Codes: L8, L82, H56

---

[1] Jakobe Bussey graduated from Duke University in May of 2024 with degrees in Computer Science, Economics (graduated with High Distinction), and a minor in Japanese. Beginning in July of 2024, he is working as a Trader for Deutsche Bank in the Fixed Income & Currencies division in New York City. He can be reached at jzbussey@gmail.com.

Acknowledgements

Firstly, I want to impart my deepest appreciation for all of the help provided by my phenomenal faculty advisor Dr. Peter Arcidiacono. Getting this entire project done within a semester was a herculean effort that could only have been achieved with a herculean professor at the helm. Your wealth of knowledge and insights made working on this project an enlightening and exhilarating experience. Secondly, I would like to thank the entire Duke Economic Analytic Lab (DEAL) research team. Economic researchers willing to utilize cutting-edge techniques and methods for the social sciences in search of the truth should be applauded. Without professors like you, papers like this would not exist.

1. Introduction

War, as a phenomenon, holds a profound impact on societies, shaping narratives, policies, and collective memory. How citizens in a nation perceive and understand wars is not only important for its historical significance, but also for shaping contemporary perspectives on current and future global conflicts. In the modern era, the role of media in disseminating information and framing narratives surrounding wars cannot be overstated. Indeed, due to the inherent dangers involved in covering conflicts, consistent news regarding wars can practically be provided only by a dedicated media apparatus, especially before the advent of ubiquitous personal cameras, providing the media with a nearly complete monopoly on initial narrative generation. Further, the portrayal of conflicts in newspapers not only reflects societal attitudes but also influences public opinion and thus provides the impetus for government action. In thinking about the power that the media has in information dissemination and narrative framing, three important questions have been brought to bear: How does the media as a whole report on war? Does the reporting differ between media outlets? Does the reporting change over time, and if so, how does it change? By answering these questions one can derive key insights into how the public appetite for war is spawned, how this appetite ebbs and flows as the state of the conflict evolves, and, perhaps, how to bring these conflicts to a quicker conclusion. To answer these crucial questions, this study turns to one of the most significant events of the 21st century - the Iraq War.

The Iraq War, spanning from 2003 to 2011, was a multifaceted conflict that garnered extensive media coverage both domestically and internationally. The portrayal of this war in the media played a critical role in shaping public perception, political discourse, and international relations (Calabrese, 2005) and thus would be an ideal candidate for answering crucial questions regarding the media and war. This paper will examine the media reporting on the Iraq War, but specifically, considering their outsized impact on other newspapers (Zhang, 2018), this paper will examine two of America’s most circulated newspapers, *The Wall Street Journal* and *The New York Times*, and they will act as a proxy for the media at large. By utilizing cutting-edge BERT (Bidirectional Encoder Representations from Transformers) models this paper will ascertain the positive, neutral, and negative sentiments of articles about the Iraq War and further will be able to discern key insights into the dynamics of how the media's views on the conflict evolve as the state of the war evolves.

This paper will employ a multistage process to perform the analysis. First, by utilizing the ProQuest database, this paper will collect articles about the Iraq War from the two leading newspapers in the United States, *The Wall Street Journal* and *The New York Times,* from January 1st, 2002, to December 31st, 2012. This paper will then apply a state-of-the-art BERT model to separate articles into groups based on their content and their relationship to certain seeding parameters. This study will then utilize a separate BERT model to perform sentiment analysis on all articles published by *The New York Times* and *The Wall Street Journal* that pertain to the Iraq War by leveraging the interpretive ability of large-language models (LLMs) to generate a positivity, negativity, and neutrality score for every article in the dataset. From there, statistical analysis is performed on *The New York Times* and *Wall Street Journal* datasets in order to discern whether or not these two large media sources report on the war homogenously. Further, a quartic regression model will be utilized to examine how negativity between the two newspapers changes over time. The findings of this section suggest that the two newspapers do report on the Iraq War differently, but they both report on it in a negative-neutral tone. Further, the results suggest that the amount of negativity does drastically change over time as the events of the war unfold. This paper then uses statistical analysis to determine if non-editorial articles report more "objectively" than their editorial counterparts, and similarly analyzes how the neutrality of the reporting changes as the war advances between editorials and non-editorials. This paper's findings suggest that non-editorials do report more objectively than editorials, and further find that while the neutrality of the editorials remains somewhat constant the objectivity of the non-editorials greatly changes based upon the events of the war. In the final section, this paper analyzes the differences in sentiment based on the topic being reported on. The paper finds that not only do positive, neutral, and negative sentiments vary greatly among the topics but they also change differently over time.

The rest of this paper is organized as follows. The next section provides information on the existing literature in the world of sentiment analysis, natural language processing, and media analysis. Section 3 will discuss the data used to explore the three key questions regarding the media's reporting on the Iraq War and offer information on the sample construction methodology. Section 4 provides the results of the experiment as well as an analysis of those results. Section 5 concludes and discusses the implications of the results found in the experiment.

2. Literature Review

This paper is positioned at the intersection of various streams of prior research and thus contributes to the literature of many fields, the most apparent being the literature that utilizes natural language processing algorithms for the analysis of text data. NLP and text analysis, more broadly, have shown great promise in economics research.

The seminal work that established text itself as a viable source for data in economics research is Gentzkow and Shapiro's "What Drives Media Slant? Evidence from U.S. Daily Newspapers" (2010), which measured the media slant of newspapers by comparing it to the language of Congressional Record text from Republicans and Democrats. This work is pivotal as it provided not only a baseline theoretical foundation of how one could rationalize firm behavior in providing news slanted towards the left or the right (this theoretical framework is expanded upon further in a later work by Gentzkow et al. (2015)), but it also pioneered the early methodology for phrase analysis, which would be important for earlier models prior to the ML (machine learning) methods developed later. Utilizing text for data extends far beyond newspaper politics. Ciliberto et al. (2019) measure airline coordination and collusive patterns using text from airline earnings calls; Hassan et al. (2019) measure the political risk faced by individual U.S. firms by examining earnings conference calls; and Baker et al. (2016) utilize newspaper text from six major

U.S. newspapers to measure economic policy uncertainty based upon newspaper coverage frequency. The fact that human readings of over 12,000 newspaper articles were required for the 2016 study showcases just how inherently expensive conducting this type of work can be. The fact that human surveys tend to be expensive, combined with the fact that surveys are liable to sampling problems due to the small samples of individuals being surveyed (Ludvigson, 2004) and that doing these human surveys typically involves a time lag between events and information gathering (i.e., one can only survey the effects of an economic downturn after the downturn, and collecting human opinions about the downturn can take a long time), provided the resolve for researchers to turn to more advanced automated solutions.

Sentiment analysis is primarily the extraction of an agent's sentiment through text, and it has become one of the fastest-growing areas of natural language processing (NLP). The impact of research in the area has flowed from the halls of academia into the world of business, as social media companies, algorithmic trading firms, and even human resource management companies have all found ways to leverage this cutting-edge technique. However, these business applications were first pioneered by economic and financial researchers. Perhaps the seminal work in the sentiment analysis space for finance, Garcia (2013) measured the financial market sentiment of the *New York Times* financial column and used it to see how that financial market sentiment affected the price of assets during recessions. Shapiro and Wilson (2022) utilized textual sentiment analysis on Federal Open Market Committee meeting transcripts from 2000–2011 in order to rigorously estimate the central bank's objective function instead of presuming that the Federal Reserve was always targeting an interest rate of 2%.

Early approaches to textual sentiment analysis, such as Garcia's paper, tended to err on "lexical" methodology, where a pre-defined list of words is assigned a score from which one can determine the positivity or negativity of a string of text. Negation, punctuation, and words like "very," "extremely," or "slightly" apply multipliers and other simple rules to further add to the context of the sentiment of sentences and thus whole phrases. However, it has been argued that “natural language is too creative and complicated—and sentiment expression too nuanced—to be fully captured by a static lexicon and a fixed list of heuristic rules.” (Shapiro et al., 2019) Thanks to advances in machine learning (ML), textual sentiment analysis using large language models (LLMs) has become the cutting-edge technique that deals with the issues of lexical methodology.

The Bidirectional Encoder Representations from Transformers (BERT) models, created by Google researchers in 2018 (Devlin et al. 2018), have become the most popular technique in machine learning text analysis. BERT models have many benefits: sentiment analysis with these BERT models can be extremely versatile and less domain-dependent than the lexical approach. Furthermore, the BERT architecture, due to its bidirectional nature, is designed to interpret the meaning of a word in the context of all other words in the sentence, rather than interpreting a word on its own or based solely on the previous word. Take, for example, the sentence: “War is good for nothing.” The lexical approach may fail to capture the full intention of the writer, as the word “good” would potentially provide a positive score, and the expression “for nothing” may not necessarily be in the lexicon as a negative expression as there are many ways one could orient “for nothing” in a positive sentence (i.e. “They volunteered all weekend **for nothing** more than the satisfaction of making a difference.”) Thus, this sentence could very well be characterized as being positive. However, by examining the entire sentence, the BERT model will be able to ascertain the author's negative sentiment toward war (given that it was trained on a large enough dataset.) As one can see in the example, Shapiro et al (2019) are correct. The challenge of creating a lexical framework that accounts for all the possible sentence combinations to determine sentiment can be rather cumbersome. Instead of focusing a great deal of time and effort on constructing a proper lexical heuristic that fits the data, one could train a BERT model on the data itself and have that BERT model learn the relationship between words in sentences and the overall sentiment on its own, which is quite powerful.

Many researchers have recognized the power that BERT models hold: In the field of finance, Sousa et al. (2019) used a BERT model to perform sentiment analysis on news articles related to the stock market, while Wang et al. (2020) analyzed negative sentiment on Chinese social media during the COVID-19 pandemic. Considering how new these machine learning models are, many areas of research are ripe for exploration.

One particularly compelling domain is the analysis of war articles, especially those related to significant historical events such as the Iraq War. Traditional sentiment analysis methods may struggle to capture the nuanced emotions and complexities expressed in such texts. Early approaches, relying on fixed lexicons and heuristic rules, often fall short of adequately deciphering the sentiment embedded within war-related narratives. However, leveraging advanced machine learning techniques, particularly BERT models, presents a promising new avenue for delving deeper into the sentiment dynamics of war articles. Thus, by performing this analysis, this study will contribute to the budding literature of ML-based sentiment analysis by creating a benchmark by which one can study sentiment dynamics in war-related narratives with LLMs like BERT.

3. Data and Sample Construction Methodology

3.1 Data

The article dataset was collected using ProQuest (PQ), an information database that archives newspaper articles from a vast array of sources spanning various topics and regions. ProQuest is renowned for its extensive collection, which encompasses reputable newspapers, scholarly journals, magazines, and other media outlets. For the study, the search function was utilized to winnow the articles of *The New York Times* and *The Wall Street Journal* to only those articles that are about the Iraq War. The following search query was used to find articles about the Iraq War from 01/01/2002 to 12/31/2012:

((((Subject(\"Iraq War\") NOT TI(Correction))AND (PUBID(X) NOT TI(\"Inside the Times\"))) NOT TI(Editor Note)) NOT DTYPE(Correction))

By choosing the PUBID for either *The Wall Street Journal* or *The New York Times*, this study is able to gather articles from either newspaper. Editors' notes and corrections were omitted as the focus of this study is on full articles. After further pruning, a total of 5,162 articles were collected from *The New York Times* with 1082 of them being editorial articles. A total of 1,497 articles were collected from *The Wall Street Journal* with 158 of the articles being editorials. Thus, there is a combined total of 6659 articles about the Iraq War between the two newspapers of interest in this dataset. Table 1 goes into more detail about the explicit breakdown of the two newspapers.

Table 1: Division of Articles

Variable Name	**Number of Articles**	**Percentage of Total**
NYT Non-Editorials	4080	61.3%
NYT Editorials	1082	16.2%
WSJ Non-Editorials	1339	20.1%
WSJ Editorials	158	2.4%
Total Number of Articles	6659	100%

Source: Author's calculations, based on ProQuest data. 6659 articles gathered over the years 2002-2012: a total of 6659 observations.

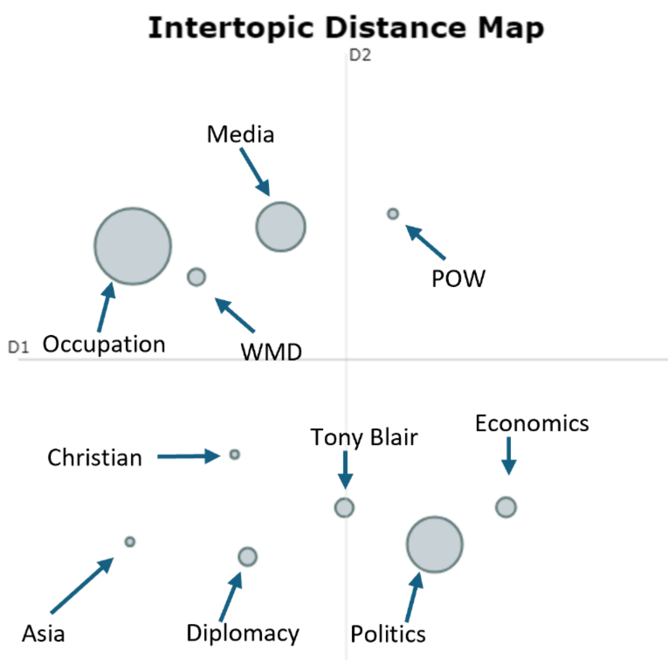
From this, we see that *The New York Times* has an outsized effect on the total dataset. When the study aggregates to create the media consensus large changes in the *New York Times* articles will be the most reflected in the total. Further, the dearth of editorial articles for *The Wall Street Journal* implies that the opinion page will have a very small impact on the overall media consensus.

3.2 Sample Construction

As previously stated, one of the primary priorities of this study is to utilize sentiment analysis to analyze articles. However, the problem with just using sentiment without topics is that without knowing the topic being discussed in the article, one cannot be sure what the writer is being positive, negative, or neutral toward. For example, an article may be ranked as negative, but there are a myriad of reasons why an article about the Iraq War can be considered negative. For example, one article could be negative because the author of the article disagrees with the methods used by the US military (making that article negative towards the US) and another article could be critical of Iraqi soldiers. Both articles would be negative, but perhaps one article is more negative about the Iraq War as a whole, and another article might be for the war and thus, anti-Iraqi military. Without the topics, one would simply presume that both articles are negative toward the war, which may not be true. But by learning the target of the sentiment (that being the topic) one can better discern and understand the positivity, negativity, and neutrality dynamics in the data. This is where BERTopic becomes an invaluable tool.

In order to see how the media changes its reporting when discussing separate issues a state-of-the-art BERT model called “BERTopic” will be utilized. The purpose of BERTopic is to generate topic representations that can be used to automatically create different topics based on the text of each of the different articles, and based upon the topics generated, sort each article into its proper topic. The study’s BERTopic model takes in the text of each of the articles, “trains” on it, and then automatically creates categories that the model calculates would best reflect the data and places each of the documents within these categories. In order to encourage the model to focus its grouping on certain topics, “seeding” can be applied (Egger & Yu, 2022). By using seeding, one can apply “seed_words” which allows certain specific words to be weighed more heavily in the topic creation process. The following set of “seed_words” were inputted into the unsupervised machine learning model: "destruction", "american", "iraq", "citizen", "wmd", "politics", and "soldier". By specifying these seeds, the model is encouraged to group based on the war itself as opposed to ulterior events surrounding the war. Initially, sixty-seven different topics were generated.[2] Figure 2 presents the Intertopic Distance map after the total number of topics is reduced to eleven.

<u>Figure 2: Intertopic Distance Map for Iraq War articles after reduction</u>



Source: Author's calculations, based on ProQuest data and BERTopic model output. 6659 articles were gathered over the years 2002-2012: a total of 11 topics were generated. (natively BERTopic does not show the largest data cluster, so only 10 are shown here)

By utilizing the Jensen-Shannon divergence method, one can measure the similarity between two probability distributions (Fugledge & Topsoe, 2004). By defining the probability distributions based on the largest data cluster, The Intertopic Distance map shows the relationship between different topic clusters as defined by their relationship with the largest data cluster. Topics that are closer together are more closely related, in that they

[2] That visualization is included in the appendix, A.3. This visualization showcases the overlap of topics in the topic space, providing ample reason to perform topic reduction.

either have many words or phrases distributed similarly to one another, and topics that are further away are more dissimilar. For example, the "Occupation" articles are more related to "WMD" articles or "POW" articles than they are to articles about "Politics" (descriptions of individual topics provided in Table 3). The larger the bubble is, the more articles (or specific words) lie under its banner, and from Figure 3 one can see that there is a great deal of concentration on three main topics and given that the largest topic is not being shown, the data is mainly defined by four main topics. Due to the fact that there is no overlap between the topics generated in Figure 3, it is implied that the topics are well-defined and well-separated. The BERTopic model, after outputting the topics, provides insight into what the actual topic is for each of the clusters by providing representative articles that define the topic (Grootendorst, 2022). For brevity, each of the topics will be given a single-word phrase that provides a general idea of what the topic cluster is primarily about. Table 3 goes into detail about each of the different topics and the number of articles under its banner.

Table 3: Topics for Iraq War articles after reduction

Topic Name	Topic Description	No. of Articles under Topic
Campaign	Articles focused primarily on the overall military campaign in Iraq. Many articles are reports on individual battles. Most general relative to the other topics.	2139
Occupation	Articles focused primarily on the Occupation of Iraq and its aftermath.	2056
Politics	Articles about the political side of the war. This includes public opinion, criticism, and electoral implications.	1084
Media	Articles about the role of media and journalists. Many of these articles constitute real-time war coverage and interviews with soldiers.	837

Economics	Article about Economics or Finance	135
Tony Blair	Articles primarily about Tony Blair, the UK, and their involvement in the Iraq War.	117
Diplomacy	Articles about diplomatic complications between France, The United Nations, The United States, and Britain.	109
WMD	Articles about WMDs, chemical weapons, and other heavy weaponry.	101
POW	Articles primarily about rescued POWs like Jessica Lynch.	33
Asia	Articles about Asian foreign policy due to the consequences of the war.	25
Christian	Articles about the Christian faith in relation to the war.	23

Source: Author's calculations, based on ProQuest data and BERTopic model output. 6659 articles gathered over the years 2002-2012: a total of 11 topics generated

It is clear that the input seeding encouraged the model to base its topics more explicitly on the war. Further analysis on each of the topics will be provided in the following section, but it is clear that in the analysis of the dynamics of how the media reported upon the war as it evolved, considering how few articles there are in the other categories, only the topics: "Campaign", "Occupation", "Politics", and "Media" should be considered.

Now that the data has been collected, a different BERT model will be used to detect the sentiment of each of the articles in the dataset. Because of the opinionated and political nature of the dataset, this study used the "Twitter-roBERTa-base for Sentiment Analysis" model. This model was trained on over 124 million tweets and for each block of text it analyzes, the model provides a positivity, neutrality, and negativity rating, the total sum of which adds up to 1 (Guzmán et al, 2023). In the case that the article is too long to be analyzed by the model natively (due to the token/word limitations of BERT models), the article is broken up into equal pieces such that the amount of text analyzed in each chunk is maximized, then an average is taken from each of its chunk ratings to determine the overall score. Each of the articles is then assigned negative, neutral, and positive sentiment scores based on the calculation of the BERT model.

4. Results

4.1 Differences between Newspapers

Table 1: Sentiment Statistics for Iraq War articles between 2002-2012

Variable Name	Average Negative, (Standard Deviation)	Average Neutral, (Standard Deviation)	Average Positive, (Standard Deviation)
All Newspapers	0.447 , (0.151)	0.484 (0.116)	0.069 (0.059)
NYT	0.450 , (0.153)	0.482 , (0.119)	0.068 , (0.058)
WSJ	0.436 , (0.145)	0.490 , (0.107)	0.074 , (0.063)

Source: Author's calculations, based on ProQuest data and BERT model output. 6659 articles gathered over the years 2002-2012: a total of 6659 observations.

Table 2: Difference in Mean Negative Sentiment Statistics Between *The New York Times* and *The Wall Street Journal* for Iraq War articles between 2002-2012

Variable Name	Negative (NYT-WSJ)
Mean	0.014
95% Confidence Interval	(0.005 , 0.022)
t-value	3.2307
p-value	0.0013
df	2544

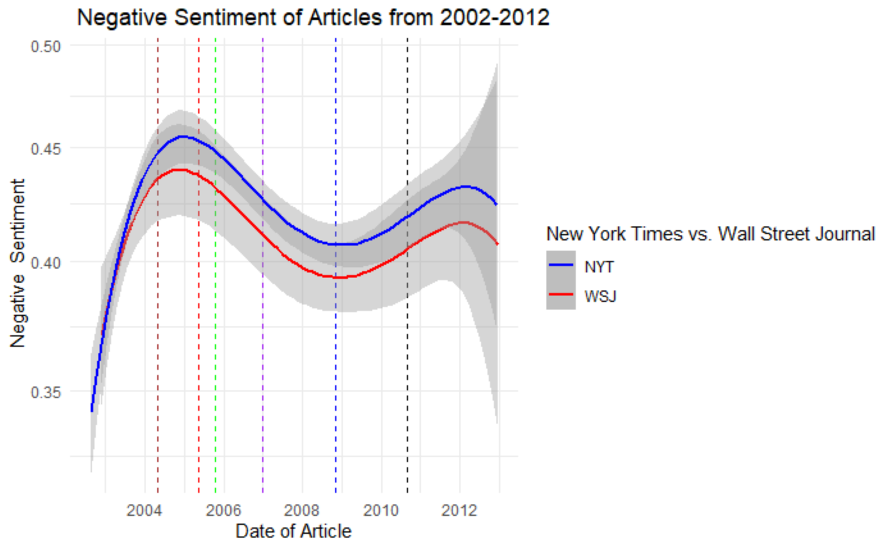
Source: Author's calculations, based on ProQuest data and BERT model output. 6659 articles gathered over the years 2002-2012: a total of 6659 observations.

As illustrated in Table 1 and further implied by Table 2, while the New York Times and the Wall Street Journal have been reporting on the same topic, The Iraq War, both newspapers reported on it differently, and this difference is statistically significant. Between 2002-2012, on average, New York Times articles had a more negative tone relative to the Wall Street Journal. Further, the difference between the average negativity of *The New York Times* and *The Wall Street Journal* is found to be statistically significant with 99% confidence. This implies that in reporting on and discussing the war, *The New York Times* was fundamentally distinct in how negative it was relative to *The Wall Street Journal.* The same can be said for both neutrality and positivity, where *The Wall Street Journal* was found to be both more neutral and more positive relative to *The New York Times* when reporting about the Iraq War.[3]

Another important aspect we can get from the data is that variation between just how positive, negative, or neutral an article is differs between newspapers. Between both *The Wall Street Journal* and *The New York Times* Table 1 shows that the most variation is found in the negative tone of the articles and the least variation is found in how positive the tone of an article is. This implies that the positivity of articles between the two newspapers is rather low and remains relatively consistent. However, as the war evolved, how negative or how neutral an article was tended to change. Indeed, as will be showcased in the dynamics section, the negativity of articles tended to oscillate as the war improved for the United States. However, it is robustly clear that over the course of the war, both newspaper outlets remained firmly negative-neutral in discussing the conflict. This baseline negative neutrality is likely the result of the fact that many of the articles discussing the war talk a great deal about the carnage the war wrought, the casualties of US soldiers, or the shaken lives of civilians.

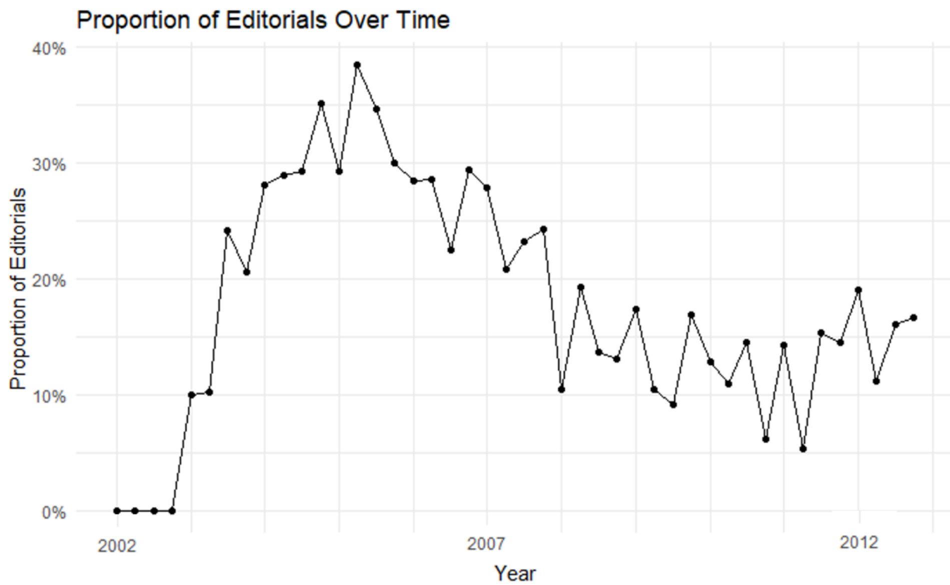
[3] The statistical significance table of the difference in terms of positivity and neutrality is provided in the appendix: Table A.2. Further commentary is also provided.

Figure 3: Negative Sentiment for Iraq War articles split by Newspaper 2002-2012



Source: Author's calculations, based on ProQuest data and BERT model output. 6659 articles gathered over the years 2002-2012: a total of 6659 observations. Brown line: (4/28/2004) Evidence of Prisoner Abuse in Abu Ghraib becomes public. Red line: (5/11/2004) Nicholas Berg's beheading video is released on the jihadist forum Muntada al-Ansar. Green line: Iraqi constitutional referendum (10/15/2005). Purple line: Execution of Saddam Hussein (12/30/2006). Blue line: Election of Barack Obama (11/4/2008). Black line: Combat mission officially ends in Iraq (9/31/2010).

<u>Figure 4: Proportion of Editorials for Iraq War articles 2002-2012</u>



Source: Author's calculations, based on ProQuest data and BERT model output. 6659 articles gathered over the years 2002-2012: a total of 6659 observations.

Figure 3 presents how that negativity changed over time between the two newspapers over the course of the war, and the results are intriguing. Firstly, the quartic regression generated for the NYT and the one generated for the WSJ are nearly identical except for just how baseline negative each newspaper is. As already stated and showcased in Tables 1 and 2, *The New York Times*, on average, is more negative than *The Wall Street Journal*, however from this Figure one can see the events that spark a change in negativity in one newspaper are the exact same ones that spark a change in the other. Based upon this Figure, one can discern that both newspapers were at their most negative in between the prisoner abuse in Abu Ghraib becoming public and the beheading of Nicholas Berg. Considering how negative those subjects are, it makes sense that the media would react rather negatively to that news. As the situation on the ground improved between 2005 and 2009 the negativity of both newspapers fell concurrently until finally, they bottomed out near the election of Barack Obama. From there they both rise concurrently until the end of the conflict.

Despite the apparent differences between the newspapers, it is clear that in the aggregate both of these newspapers react to events similarly. This is most likely due to the fact that non-editorial articles (which make up the bulk of the dataset) from both newspapers would presumably be reporting on the same event and thus would most likely use similar language to report on it. Further, Figure 4 indicates that the high amount of negativity between 2004-2006 might be due to the fact that a large share of articles during that period were editorials. As will be discussed, in the next section these mostly negative editorials further bring down the entire newspaper aggregate.

4.2 Differences between Editorials and Non-Editorials

Table 5: Sentiment Statistics for Iraq War articles between 2002-2012 divided by the type of article (Editorial vs. Non Editorial)

Variable Name	Average Negative, (Standard Deviation)	Average Neutral, (Standard Deviation)	Average Positive, (Standard Deviation)
All Newspapers (Only Editorials)	0.542 , (0.148)	0.397 , (0.110)	0.061 , (0.052)
All Newspapers (Non-Editorials)	0.425 , (0.144)	0.504 , (0.108)	0.071 , (0.061)
NYT (Only Editorials)	0.551 , (0.145)	0.390 , (0.109)	0.060 , (0.050)
NYT (Non-Editorials)	0.423 , (0.144)	0.506 , (0.109)	0.070 , (0.060)
WSJ (Only Editorials)	0.481 , (0.148)	0.445 , (0.102)	0.075 , (0.060)
WSJ (Non-Editorials)	0.431 , (0.144)	0.495 , (0.106)	0.074 , (0.063)

Source: Author's calculations, based on ProQuest data and BERT model output. 6659 articles gathered over the years 2002-2012: a total of 6659 observations.

Table 6: Difference in Mean Neutral Sentiment Greater Than 0 Statistics Between Non-editorials and Editorials for Iraq War articles between 2002-2012

Variable Name	Neutral (All Non Ed. – All Only Ed.) > 0	Neutral (NYT Non Ed. – NYT Only Ed.) > 0	Neutral (WSJ Non Ed. – WSJ Only Ed.) > 0
Mean	0.107	0.116	0.050
95% Confidence Interval	(0.101 , Inf.)	(0.029 , Inf.)	(0.036 , Inf.)
t-value	30.897	31.154	5.814
p-value	< 2.2e-16	< 2.2e-16	1.041e-8
df	1829	1693	199

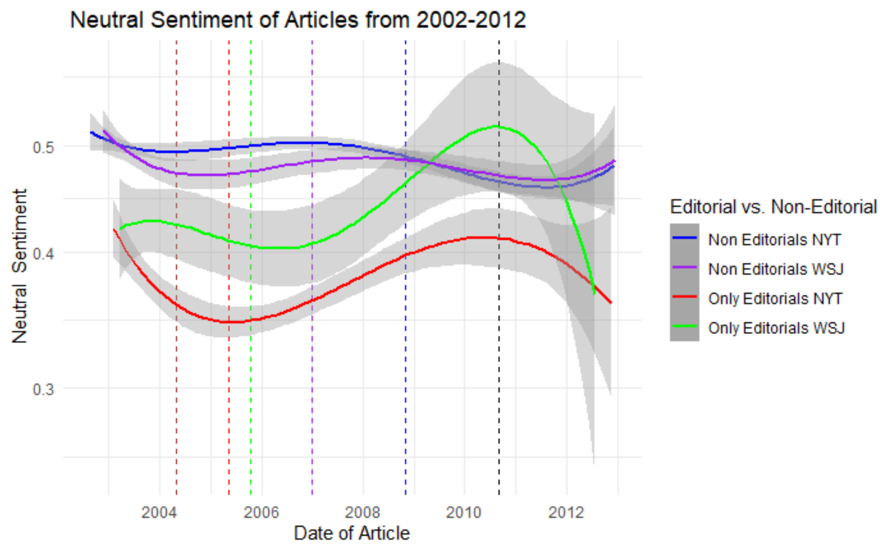
Source: Author's calculations, based on ProQuest data and BERT model output. 6659 articles gathered over the years 2002-2012: a total of 6659 observations.

Table 5 showcases the clear differences in how the Iraq War was discussed between the editorials and typical reporting. As one would presuppose, the average neutrality (or objectivity) of articles is higher among non-editorials relative to editorial articles. This relationship is shown to be statistically significant based upon Table 6, which shows that in each category (between *The New York Times* editorial articles and non-editorial articles, *The Wall Street Journal* editorial articles and non-editorial articles, and in aggregating all editorial articles against all non-editorial articles) the non-editorial articles are always found to be more neutral (or objective) than their editorial counterparts with over 99% confidence. This most likely stems from the fact that editorial articles tend to be more opinionated than their typical article counterparts and thus are less objective. It is interesting to note that of the two newspapers, the *Wall Street Journal's* editorial articles are closer in terms of "objectivity" to their non-editorial articles than *The Wall Street Journal's* editorial articles are to non-editorial articles.

However, in aggregating both *New York Times* and *Wall Street Journal* articles together, the largest difference in means is not found between average neutrality in editorials vs. non-editorials. Instead, it is clear that the largest difference between editorials discussing the Iraq War and non-editorials is actually found in how negative they are. Since the difference in positivity between editorials and non-editorials is marginal, it is evident that should a writer wish to express an opinion or a belief about the Iraq War that is distinct from the reporting, that belief is most likely to be a negative one. This is most likely true because of just how many editorial articles there are about the Iraq War that come from *The New York Times.* Considering that *The New York Times'* editorial articles are the most negative ones found in the dataset, they tend to crowd out the more positive ones (positive relative to *The New York Times)* found in the editorial articles of *The Wall Street Journal*. Nevertheless, the amount of true positivity exhibited by either newspaper, no matter which type of article one examines, is rather small.

Further, the variance of positivity, negativity, and neutrality between typical editorial articles and non-editorial articles tends to differ slightly. In aggregation, editorials have a higher variance when it comes to negativity and neutrality, but this relationship is flipped when it comes to positivity, whereas non-editorials tend to have more variability in positivity. The variance in positivity is again most likely due to the influence of *The New York Times*, which boasts a much smaller positive variability relative to *The Wall Street Journal* implying that positive sentiments or opinions regarding the war were more uniform among *New York Times* editorial writers relative to *The Wall Street Journal*.

<u>Figure 7: Neutral Sentiment for Iraq War articles split by Editorials and Newspaper between 2002-2012</u>



Source: Author's calculations, based on ProQuest data and BERT model output. 6659 articles gathered over the years 2002-2012: a total of 6659 observations. Brown line: (4/28/2004) Evidence of Prisoner Abuse in Abu Ghraib becomes public. Red line: (5/11/2004) Nicholas Berg's beheading video is released on the jihadist forum Muntada al-Ansar. Green line: Iraqi constitutional referendum (10/15/2005). Purple line: Execution of Saddam Hussein (12/30/2006). Blue line: Election of Barack Obama (11/4/2008). Black line: Combat mission officially ends in Iraq (9/31/2010).

Figure 7 reinforces the interpretation of the variance found in Table 6, but also provides some interesting conclusions in regard to when the objectivity of the outlets changed. The non-editorial articles tend to keep in line with one another, only marginally differing in objectivity from each other over the course of the Iraq War. *The New York Times* remained more objective relative to *The Wall Street Journal* until the election of Barack Obama in 2008. From then on, the neutrality of *The Wall Street Journal* supersedes it, but again this difference is marginal. However, the differences in objectivity between *The Wall Street Journal* and *The New York Times* in terms of their editorials are rather stark. For nearly the entire dataset, the editorials from *The Wall Street Journal* are more objective than the ones from *The New York Times*. However, there is a small period near the end of the dataset where the opposite is true. While the non-editorials hardly reacted to the changing landscape of the war, editorials were quite attuned to them. *The New York Times* editorials became the least objective near the beheading of Nicolas Berg, while *The Wall Street Journal* was at its least objective in between the Iraqi constitutional referendum and the execution of Saddam Hussien. After these periods there is a sharp rise in objectivity, up until the combat mission official ends, at which point there is a concurrent sharp drop in objectivity, especially from *The Wall Street Journal.* This primarily has to do with the fact that articles after the war concluded were quite negative in their reflections on the conflict between both newspapers. This data seems to suggest that during a crisis editorials are much more willing to swing to the emotion of the times and may more accurately reflect the opinions of the populace.

4.3 Differences between Topics

Table 8: Sentiment Statistics for Iraq War articles between 2002-2012 divided by the topic of the article

	Average Negative, (Standard Deviation)	Average Neutral, (Standard Deviation)	Average Positive, (Standard Deviation)	%Share of Articles
All Newspapers Topic: All	0.447 , (0.151)	0.484 , (0.116)	0.069 , (0.059)	100.0 %
NYT Topic: All	0.450 , (0.153)	0.482 , (0.119)	0.068 , (0.058)	77.5 %
WSJ Topic: All	0.436 , (0.145)	0.490 , (0.107)	0.074 , (0.063)	22.5 %
All Newspapers Topic: Campaign	0.452 , (0.148)	0.480 , (0.117)	0.068 , (0.055)	32.1 %
NYT Topic: Campaign	0.454 , (0.152)	0.479 , (0.120)	0.067 , (0.055)	79.2 %
WSJ Topic: Campaign	0.442 , (0.135)	0.487 , (0.104)	0.071 , (0.054)	20.8 %
All Newspapers Topic: Occupation	0.476 , (0.154)	0.469 , (0.123)	0.055 , (0.050)	30.9 %
NYT Topic: Occupation	0.483 , (0.155)	0.465 , (0.126)	0.052 , (0.046)	73.1 %
WSJ Topic: Occupation	0.459 , (0.151)	0.478 , (0.115)	0.062 , (0.059)	26.9 %

All Newspapers Topic: Politics	0.426 , (0.136)	0.504 , (0.107)	0.070 , (0.049)	16.3 %
NYT Topic: Politics	0.432 , (0.136)	0.502 , (0.111)	0.066 , (0.043)	78.9 %
WSJ Topic: Politics	0.406 , (0.133)	0.509 , (0.092)	0.085 , (0.063)	21.1 %
All Newspapers Topic: Media	0.420 , (0.161)	0.485 , (0.111)	0.095 , (0.077)	12.6 %
NYT Topic: Media	0.423 , (0.161)	0.484 , (0.111)	0.094 , (0.075)	87.0 %
WSJ Topic: Media	0.401 , (0.160)	0.494 , (0.106)	0.105 , (0.085)	13.0 %
All Newspapers Topic: Economics	0.370 , (0.145)	0.510 , (0.099)	0.120 , (0.077)	2.0 %
NYT Topic: Economics	0.402 , (0.146)	0.490 , (0.102)	0.108 , (0.067)	55.6 %
WSJ Topic: Economics	0.330 , (0.133)	0.536 , (0.089)	0.135 , (0.087)	44.4 %
All Newspapers Topic: Tony Blair	0.439 , (0.100)	0.492 , (0.073)	0.069 , (0.053)	1.8 %
NYT Topic: Tony Blair	0.436 , (0.101)	0.494 , (0.073)	0.069 , (0.056)	79.5 %
WSJ Topic: Tony Blair	0.449 , (0.098)	0.485 , (0.072)	0.066 , (0.040)	20.5 %
All Newspapers Topic: Diplomacy	0.403 , (0.140)	0.520 , (0.106)	0.077 , (0.063)	1.6 %
NYT Topic: Diplomacy	0.393 , (0.144)	0.525 , (0.104)	0.082 , (0.072)	66.1 %
WSJ Topic: Diplomacy	0.423 , (0.131)	0.509 , (0.109)	0.067 , (0.038)	33.9 %
All Newspapers Topic: WMDs	0.438 , (0.146)	0.499 , (0.114)	0.063 , (0.058)	1.5 %
NYT Topic: WMDs	0.421 , (0.150)	0.511 , (0.115)	0.069 , (0.063)	79.2 %
WSJ Topic: WMDs	0.506 , (0.109)	0.454 , (0.098)	0.041 , (0.017)	20.8 %
All Newspapers Topic: POWs	0.293 , (0.149)	0.577 , (0.121)	0.130 , (0.148)	0.5 %
NYT Topic: POWs	0.287 , (0.148)	0.581 , (0.121)	0.132 , (0.150)	97.0 %

WSJ Topic: POWs	0.473 , (NA)	0.446 , (NA)	0.081 , (NA)	3.0 %
All Newspapers Topic: Asia	0.447 , (0.122)	0.495 , (0.105)	0.058 , (0.037)	0.4 %
NYT Topic: Asia	0.416 , (0.104)	0.510 , (0.097)	0.074 , (0.046)	40.0 %
WSJ Topic: Asia	0.467 , (0.133)	0.485 , (0.113)	0.048 , (0.026)	60.0 %
All Newspapers Topic: Christian	0.316 , (0.147)	0.557 , (0.087)	0.127 , (0.108)	0.3 %
NYT Topic: Christian	0.288 , (0.146)	0.570 , (0.088)	0.141 , (0.114)	82.6 %
WSJ Topic: Christian	0.445 , (0.060)	0.495 , (0.052)	0.060 , (0.015)	17.4 %

Source: Author’s calculations, based on ProQuest data and BERT model output. Topic division based upon SpanBERT model output. 6659 articles gathered over the years 2002-2012: a total of 6659 observations.

As presented in Table 8, there is quite a distinction in the sentiment of how different topics were discussed during the war. Overall, the topic that received the most negative sentiment was “Occupation” that being articles surrounding the criticism of Paul Bremer’s disbanding of the Iraqi army and the struggles that came along with occupying Iraq after Saddam Hussein’s regime was toppled in 2003. Further, among all articles, the next most negative topic was “Campaign”, that being articles primarily about the military campaign in Iraq itself. In both of these cases and especially in the former case, *The New York Times* is much more negative on these subjects on average relative to *The Wall Street Journal*. Indeed, for *The New York Times*, both of these topics receive the most negativity on average relative to the other topics. Because these two topics constitute the majority of the dataset, *The New York Times* has an overall strong negative score on the dataset as a whole. Considering that the “Campaign” and the “Occupation” account for ~63% of the data, this negativity brings up the overall negativity of the data.

However, that does not necessarily imply that *The Wall Street Journal* is free from negativity. In fact, the topic that received the most negativity on average (out of individual newspapers, not the aggregate most negative) was from *The Wall Street Journal* in its articles about “WMDs” and chemical weaponry. This has to do with the fact that articles discussing chemical weaponry typically discuss them in an extremely negative light and further, such articles might bring up explicit details of the disfigurements incurred due to chemical attacks. In addition to how negative those aspects are, the negativity is further exacerbated by the fact that there were not in fact “WMDs” discovered, providing ample ammunition to criticize the Bush Administration and thus create articles with more negative tones. However, the barrels were not aimed squarely at the Bush Administration, many *Wall Street Journal* writers were also quite critical of the United Nations as well as Sadam himself, adding more areas where articles on this topic could be quite negative.

On the more positive side, Table 8 has interesting conclusions regarding which topics were the least negative and or most positive. Among all articles as a whole, the articles with, on average, the highest positive rating and further the lowest negativity rating were articles about rescued POWs like Jessica Lynch. As such, there isn’t too much to be negative about besides the treatment they endured or the crisis that befell them such that they were captured and thus there is a much lower negative rating for such articles. Indeed, some of the most positive articles in the entire dataset are about the rescued POWs. The next highest positive rating was for articles relating to Christianity and it goes to the power that the “Christian Right” held at that time (Williams, 2012.) Many of these articles made reference to Christians praying about the war or the current Pope’s reaction to the war. A majority of the articles that were discussing the Christians at home or abroad were non-editorials, so the room to be positive was limited, which is one of the reasons why the “Christianity” topic has one of the highest neutral ratings in the entire dataset. Further, of the articles that were labeled as “Christian” and were editorials, not a single one was from the WSJ, which would explain why the NYT is so much more positive relative to the WSJ.

Table 9: Proportion of Topics on aggregate for Iraq War articles between 2002-2012 Newspaper and Editorials

	% Share of Campaign	% Share of Occupation	% Share of Politics	% Share of Media	% Share of Economics
NYT (Only Editorials)	20.1 %	12.9 %	20.9 %	11.7 %	8.1 %
NYT (Non-Editorials)	59.2 %	60.2 %	57.9 %	75.3 %	47.4 %
WSJ (Only Editorials)	2.1 %	2.5 %	3.7 %	1.0 %	0.7 %
WSJ (Non-Editorials)	18.7 %	24.4 %	17.4 %	12.1 %	43.7 %

	% Share of Tony Blair	% Share of Diplomacy	% Share of WMDs	% Share of POWs	% Share of Asia	% Share of Christian
NYT (Only Editorials)	12.8 %	11.0 %	13.9 %	9.1 %	12.0 %	17.4 %
NYT (Non-Editorials)	66.7 %	55.0 %	65.3 %	87.9 %	28.0 %	65.2 %
WSJ (Only Editorials)	2.6 %	5.5 %	1.0 %	3.0 %	8.0 %	0.0 %
WSJ (Non-Editorials)	17.9 %	28.4 %	19.8 %	0.0 %	52.0 %	17.4 %

Source: Author's calculations, based on ProQuest data and BERT model output. Topic division based upon SpanBERT model output. 6659 articles gathered over the years 2002-2012: a total of 6659 observations. Note: This is the percent share for each of the newspapers and editorials on each topic. Meaning 59.2% of "Campaign" articles were NYT Non-Editorials.

Table 9 presents the overall impact that each newspaper had on each topic based on the aggregate number of newspapers on a topic. Because of the massive disparity in the total number of articles between *The New York Times* and *The Wall Street Journal*, *The New York Times* dominates the impact on the vast majority of sections editorially and non-editorially. In fact, the only time where the impact on a topic is larger for *The Wall Street*

Journal relative to *The New York Times* is in the "Asia" topic, which only accounts for 0.4% of the dataset.

Table 10: Proportion of Topics in each Newspaper for Iraq War articles between 2002-2012 Newspaper and Editorials

	% Share of Campaign	% Share of Occupation	% Share of Politics	% Share of Media	% Share of Economics
NYT (Only Editorials)	8.3%	5.2%	4.4%	1.9%	0.2%
NYT (Non-Editorials)	24.5%	24.0%	12.2%	12.2%	1.2%
WSJ (Only Editorials)	3.0%	3.4%	2.7%	0.5%	0.1%
WSJ (Non-Editorials)	26.7%	33.5%	12.6%	6.7%	3.9%

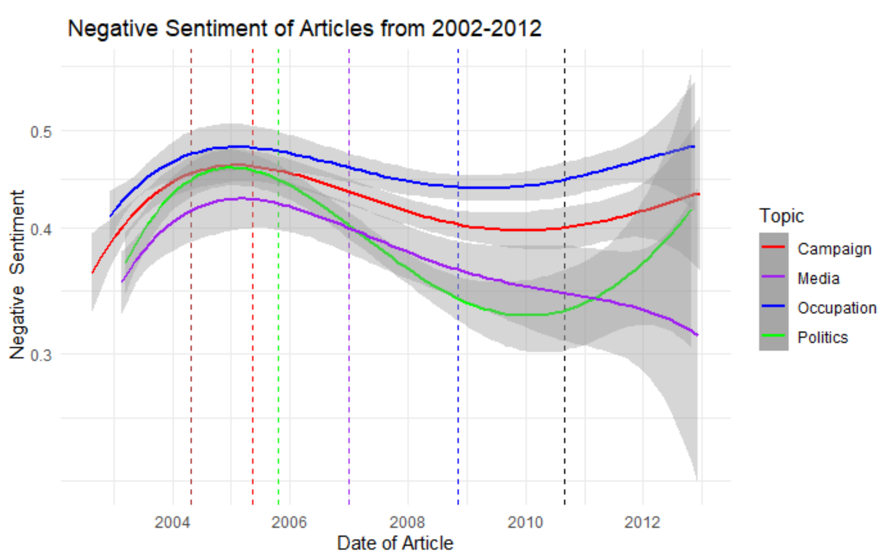
	% Share of Tony Blair	% Share of Diplomacy	% Share of WMDs	% Share of POWs	% Share of Asia	% Share of Christian
NYT (Only Editorials)	0.3%	0.2%	0.3%	0.1%	0.1%	0.1%
NYT (Non-Editorials)	1.5%	1.2%	1.3%	0.6%	0.1%	0.3%
WSJ (Only Editorials)	0.2%	0.4%	0.1%	0.1%	0.1%	0.0%
WSJ (Non-Editorials)	1.4%	2.1%	1.3%	0.0%	0.9%	0.3%

Source: Author's calculations, based on ProQuest data and BERT model output. Topic division based upon SpanBERT model output. 6659 articles gathered over the years 2002-2012: a total of 6659 observations. Note: This is the percentage of each newspaper devoted to each topic and editorial. Meaning 24.5% of the NYT articles from 2002-2012 about the Iraq War were non-editorial articles on the topic "Campaign."

Table 10 indicates where each newspaper expended its resources, and further what their readership is interested in reading about. Being that there is a finite amount of space in every newspaper, each newspaper has to divide the newspaper into different pieces devoted to different topics and further delegate stories to non-editorial reporters or editorial writers. Table 10 shows that *The Wall Street Journal* expended a similar amount of resources in reporting on the topic "Campaign", "Politics", and "WMDs" as *The New York Times* did. However, *The Wall Street Journal* committed more resources to the "Occupation" relative to *The New York Times*, and vice-versa for the "Media" topic. Further, as expected, *The Wall Street Journal* committed more resources to "Economics" relative to *The New York Times* but in terms of aggregate impact, Table 9 indicates that *The New York Times* actually had a larger impact in that space.

Tables 8-10 showcase how the war was covered, differently and similarly, between *The Wall Street Journal* and *The New York Times* between 2002-2012 among topics. However, over the course of the decade, the state of play for the United States in Iraq drastically changed. From the initial surge to the occupation, and finally to the repatriation, the news changed as the situation on the ground changed.

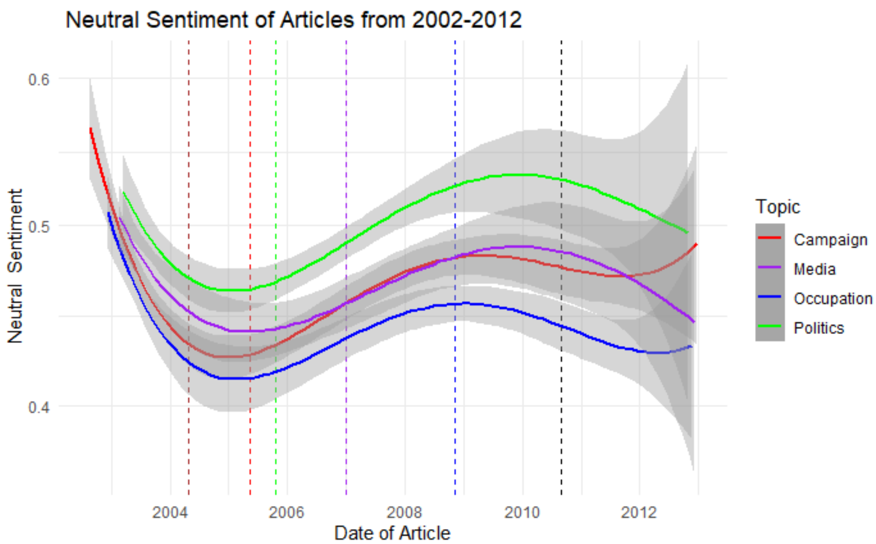
Figure 11: Negative Sentiment for Iraq War articles split by Topic between 2002-2012



Source: Author's calculations, based on ProQuest data and BERT model output. Topic division based upon SpanBERT model output. 6659 articles gathered over the years 2002-2012: a total of 6659 observations. Brown line: (4/28/2004) Evidence of Prisoner Abuse in

Abu Ghraib becomes public. Red line: (5/11/2004) Nicholas Berg's beheading video is released on the jihadist forum Muntada al-Ansar. Green line: Iraqi constitutional referendum (10/15/2005). Purple line: Execution of Saddam Hussein (12/30/2006). Blue line: Election of Barack Obama (11/4/2008). Black line: Combat mission officially ends in Iraq (9/31/2010).

Figure 12: Neutral Sentiment for Iraq War articles split by Topic between 2002-2012



Source: Author's calculations, based on ProQuest data and BERT model output. Topic division based upon SpanBERT model output. 6659 articles gathered over the years 2002-2012: a total of 6659 observations. Brown line: (4/28/2004) Evidence of Prisoner Abuse in Abu Ghraib becomes public. Red line: (5/11/2004) Nicholas Berg's beheading video is released on the jihadist forum Muntada al-Ansar. Green line: Iraqi constitutional referendum (10/15/2005). Purple line: Execution of Saddam Hussein (12/30/2006). Blue line: Election of Barack Obama (11/4/2008). Black line: Combat mission officially ends in Iraq (9/31/2010).

Figures 11 and 12 provide fascinating insights into how different topics during the Iraq War changed in concert with one another. For example, from Figure 12 the quartic model would suggest that neutral sentiment fluctuations between 2002 and 2012 are quite related. The sentiment of political articles tended to be the most neutral, most likely due to the amount of non-editorial articles that account for a large proportion of the articles under the “Politics” label, and the articles under the “Occupation” tend to be the least neutral, most likely owing to the larger proportion of editorials, however, this neutrality decreases in tandem for all four of the variables. Each of these functions reaches its minimum, when they are the least neutral, in between the time when evidence of prisoner abuse in Abu Ghraib became public and when the video of Nicolas Berg’s beheading was publicly released. Understandably, each of these events has relationships to each of these variables so naturally they should shift, however, the uniformity of the shift is particularly interesting. The strongly worded editorial articles that surfaced after these events perhaps would decrease the “objectivity” of the more neutral non-editorial articles. Further, these editorial articles reacting to the events crowded out a lot of non-editorial articles. From there each graph rises concurrently with one another on positive news from the American perspective. They each continue to rise concurrently until around when American troops are returned home. From there they fray a little, further showcased by the larger standard error in the reflections about the war. The reason for this fraying, and why the standard error increased in 2012 is that in 2012 the war in Iraq was no longer a primary topic of interest, and as such articles about the war dwindled in that later period.

Figure 11 follows in the tradition of Figure 12 presenting a striking inflection point between the prisoner abuse in Abu Ghraib and the Berg beheading. All four lines characterize that period as the most negative in terms of sentiment. Further, all of the topics do tend to decrease their negativity after the Iraqi election in 2005, however, the rate by which they are decreasing is different for each of the lines with “Politics” reacting the most aggressively to the better conditions of the war. Further, once the war ends, “Politics” returns back to the state it was near its inflection point. This could perhaps be best explained by the midterm season kicking off in full force near the overall end of the war in Iraq. The “Occupation” and “Campaign” curves follow in “Politics’” footsteps, each reacting similarly as time changes but distinctly in that the massive variation between 2002-2012 for “Politics” was far more subdued among “Occupation” and “Campaign.” However, the “Media” topic does not seem to follow the “Politics” trend. While up to the beheading they were raised in a concurrent fashion, the “Media” topic after that period then diverges from the others and continues to become less and less negative. Since the “Media” topic primarily covers articles where journalists are either on the battlefield or when they are having conversations with soldiers or civilians (American or Iraqi) there are quite a few opportunities to be negative near the beginning. At the start of the war, war correspondents are in the middle of the fighting, and in describing what they are seeing, the articles can be rather negative. However, as the situation improved, as the amount of fighting decreased, and as it looked like American soldiers would return home soon, there were fewer areas for these more negative articles to appear. As this negativity decreases, it is then transferred into both positivity and neutrality, which is one of the reasons why both of those sentiments increase during that latter period. From this, for the war in Iraq, it is clear that articles under the “Media” banner are more explicitly linked to the state of the war itself than the other variables. [4]

Every war is different and that goes the same for how the media covers it as well, however, this analysis can provide general insight into patterns of how media narratives transform and differ between outlets. Further, this work showcases general patterns for certain topics of data that may hold true for other conflicts as well.

5. Conclusion

This study investigates three central questions regarding the media and the Iraq War: How did the media, that being *The Wall Street Journal* and *The New York Times*, report about the Iraq War? Does the reporting differ between *The Wall Street Journal* and *The New York Times* regarding the Iraq War? Does the reporting about the Iraq War change over time, and if so, how does it change?

Based on the results of the study, it is clear that, on average, media reporting on the Iraq War was very negative-neutral, with most of the neutrality arising from the non-editorials and the negativity coming from editorials. This is further exemplified by the fact

[4] An analysis of positive sentiment dynamics on topics is included in the appendix: A.1

that even when discussing topics where there is a great deal of room for positivity like rescued POWs, both newspapers funneled a lot of typical negativity into neutrality instead of positivity. The results of the difference in mean sentiment statistics provide strong evidence for a difference in reporting between *The New York Times* and *The Wall Street Journal*. The robustness of the results follows from the fact that the differences were found to be statistically significant for negativity with 99% confidence. Further evidence is apparent when comparing the differences between editorials and non-editorials for the two newspapers, as the distinctions became even more stark between what each newspaper chose to report on and with what tone they reported on it. Thus, one can surmise that although *The Wall Street Journal* and *The New York Times* were reporting on the same topic, and even though they fluctuated to new events similarly, the way the Iraq War was reported between the two newspapers was distinct. This study also confirms that indeed the way the media discusses the Iraq War has evolved as the war evolved. Further, these changes can differ based on the topic being discussed in the articles. Based upon the results of this study, the relationship between the media and war is not static, the way the media discusses war is intrinsically linked to the state of the war on the ground.

While the results of this paper are quite robust for the *Wall Street Journal* and *The New York Times* regarding the Iraq War, future research could expand on this topic and see if the results found in this paper apply to other wars in the 20th and the 21st century. Further, even in studying the Iraq war, future researchers could expand the media definition to include other large newspapers like *The Washington Post*. It would also be beneficial to examine how newspapers in other countries react to conflicts that don’t involve that newspaper’s country, would those newspapers be more objective? Another interesting area of additional study would be to examine polling data among Americans during the Iraq War and compare it to the positive sentiment espoused by newspapers. Are they related, and if so, does a change in the media perspective happen before a change in the public perspective or vice-versa? Answering these questions would provide even deeper insight into the relationship between the media and war.

List of References

Baker, S. R., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. The quarterly journal of economics, 131(4), 1593-1636.

Calabrese, A. (2005). Casus Belli: US media and the justification of the Iraq War. Television & New Media, 6(2), 153-175.

Ciliberto, F., Aryal, G., & Leyden, B. (2019). Coordinated capacity reductions and public communication in the airline industry. Becker Friedman Institute wp 2018, 11.

Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).

Egger, R., & Yu, J. (2022). A topic modeling comparison between lda, nmf, top2vec, and bertopic to demystify Twitter posts. Frontiers in sociology, 7, 886498.

Fuglede, B., & Topsoe, F. (2004, June). Jensen-Shannon divergence and Hilbert space embedding. In International symposium on information theory, 2004. ISIT 2004. Proceedings. (p. 31). IEEE.

Garcia, D. (2013). Sentiment during recessions. The journal of finance, 68(3), 1267-1300.

Gentzkow, M., & Shapiro, J. M. (2010). What drives media slant? Evidence from US daily newspapers. Econometrica, 78(1), 35-71.

Gentzkow, M., Shapiro, J. M., & Stone, D. F. (2015). Media bias in the marketplace: Theory. In Handbook of media economics (Vol. 1, pp. 623-645). North-Holland.

Grootendorst, M. (2022). BERTopic: Neural Topic Modeling With a Class-Based TF-IDF Procedure. arXiv:2203.05794v0571. Available online at: https://arxiv.org/pdf/2203.05794.pdf.

Guzmán, S. E. V., Alatorre, G. D. L. R., Cantu-Ortiz, F. J., & Ceballos Cancino, H. G. (2023, February). Holistic Tweet-Based Sentiment Analysis on the China-Taiwan Conflict. In International Congress on Information and Communication Technology (pp. 435-446). Singapore: Springer Nature Singapore.

Hassan, T. A., Hollander, S., Van Lent, L., & Tahoun, A. (2019). Firm-level political risk: Measurement and effects. The Quarterly Journal of Economics, 134(4), 2135-2202.

Ludvigson, S. C. (2004). Consumer confidence and consumer spending. Journal of Economic perspectives, 18(2), 29-50.

Shapiro, A. H., & Wilson, D. J. (2022). Taking the fed at its word: A new approach to estimating central bank objectives using text analysis. The Review of Economic Studies, 89(5), 2768-2805.

Shapiro, A. H., Sudhof, M., & Wilson, D. J. (2022). Measuring news sentiment. Journal of econometrics, 228(2), 221-243.

Sousa, Matheus Gomes, et al. (2019). BERT for stock market sentiment analysis. IEEE 31st international conference on tools with artificial intelligence (ICTAI). IEEE, 2019.

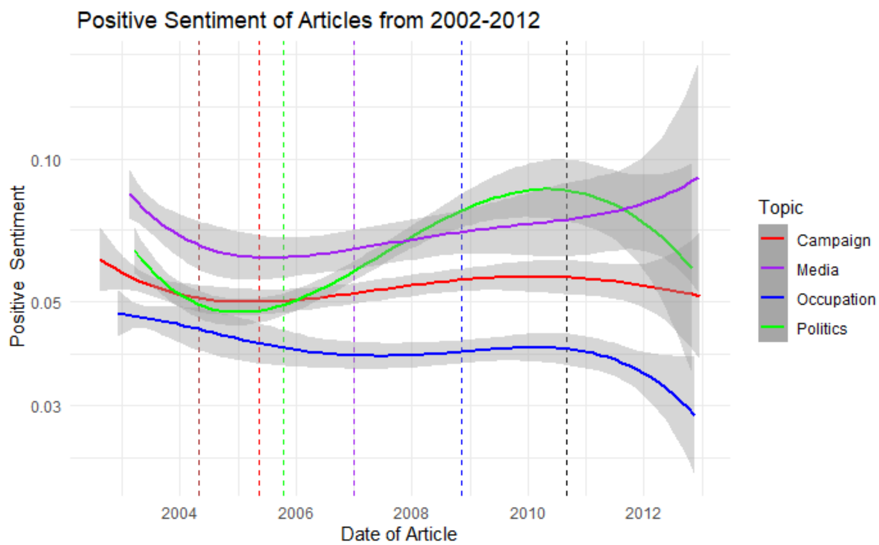
Wang, T., Lu, K., Chow, K. P., & Zhu, Q. (2020). COVID-19 sensing: negative sentiment analysis on social media in China via BERT model. Ieee Access, 8, 138162-138169.

Williams, D. K. (2012). God's own party: The making of the Christian right. Oxford University Press, USA.

Zhang, X. (2018). Intermedia agenda-setting effect in corporate news: Examining the influence of The New York Times and The Wall Street Journal on local newspapers. Journal of Applied Journalism & Media Studies, 7(2), 245-263.

Appendix

Figure A.1: Positive Sentiment for Iraq War articles by Topic between 2002-2012



Source: Author's calculations, based on ProQuest data and BERT model output. Topic division based upon SpanBERT model output. 6659 articles gathered over the years 2002-2012: a total of 6659 observations. Brown line: (4/28/2004) Evidence of Prisoner Abuse in Abu Ghraib becomes public. Red line: (5/11/2004) Nicholas Berg's beheading video is released on the jihadist forum Muntada al-Ansar. Green line: Iraqi constitutional referendum (10/15/2005). Purple line: Execution of Saddam Hussein (12/30/2006). Blue line: Election of Barack Obama (11/4/2008). Black line: Combat mission officially ends in Iraq (9/31/2010).

Figure A.1 showcases a much more subdued inflection point for positive sentiment, "Media", "Politics", and "Campaign" all exhibit some form of inflection while articles about the "Occupation" hardly shift and remain on the same course. One of the primary reasons why positivity was not shaken nearly as much as the other sentiments is because there wasn't much positivity, to begin with, certainly not in the beginning periods of the conflict. Figure A.1 shows that divergence between the different topics is highest in the realm of positive sentiment. While "Media" and "Politics" track with each other, they tend to diverge after the 2008 election of Barack Obama. While the situation in Iraq improving intrinsically improves "Media" (as discussed previously), the reflections upon the military campaign in 2012 after repatriation provide more opportunity for negative reflection on the war, and thus there is a mild decrease in positivity. The extreme variation in "Politics" is also seen in positivity as it was in negativity and neutrality. Since the soldiers were repatriated, there would be increased positive sentiment in the American political sphere and thus the data shows that increase. The articles about the "Occupation" seem to remain relatively constant in positivity throughout the occupation. However, after American soldiers returned home, the 2012 reflection upon the war provided even more ammunition to talk negatively or more objectively about the occupation and thus there was a strong decline in positivity by the end of 2012.

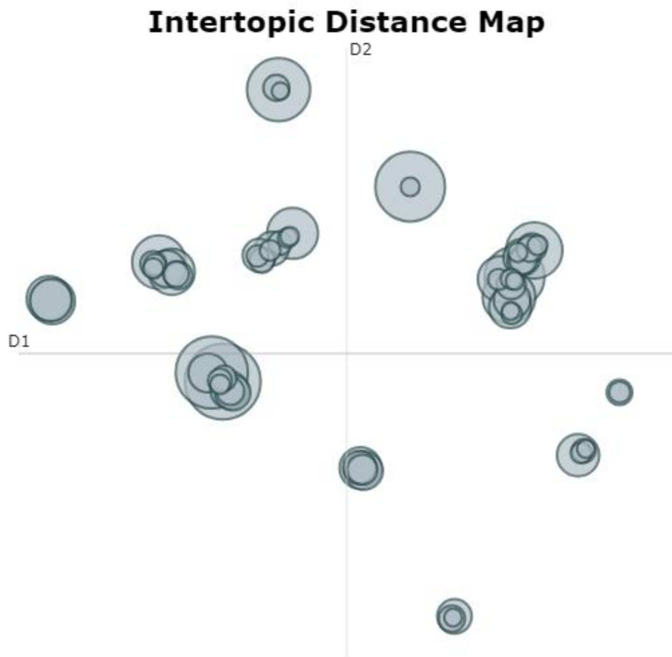
Table A.2: Difference in Mean Neutral and Positive Sentiment Statistics Between *The New York Times* and *The Wall Street Journal* for Iraq War articles between 2002-2012

Variable Name	Neutral (NYT-WSJ)	Positive (NYT-WSJ)
Mean	-0.008	-0.006
95% Confidence Interval	(-0.010 , -0.003)	(-0.014 , -0.002)
t-value	-2.4387	-3.351
p-value	0.0148	0.0008
df	2669	2287

Source: Author's calculations, based on ProQuest data and BERT model output. 6659 articles gathered over the years 2002-2012: a total of 6659 observations.

Similar to their differences in negativity, their differences in neutrality and positivity were also statistically significant, however, their difference in positivity was more statistically significant with 99% confidence versus the difference in neutrality which only had 95% confidence (p-value of 0.0148).

Figure A.3: Intertopic Distance Map for Iraq War articles before reduction



Source: Author's calculations, based on ProQuest data and BERTopic model output. 6659 articles gathered over the years 2002-2012: a total of 67 topics generated (66 shown.)