Troublemakers in the Streets?

A Framing Analysis of Newspaper Coverage of Protests in the UK 1992-2017

Johannes B. Gruber, MSc, MA

Submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

School of Social and Political Sciences

College of Social Sciences

University of Glasgow

June 2021

Abstract

The main objective of this thesis is to contribute to a more systematic understanding of how mainstream news media in liberal democracies report about protests. Existing research indicates that when mainstream news media report about demonstrations, protesters often face delegitimising coverage. This phenomenon, known as the "(journalistic) protest paradigm", is thought to be a default mindset that leads journalists to emphasize the method of protesters over their message — restricting the impact of one of few tools citizens have to raise important issues.

More recent studies, however, suggest a more mixed picture, indicating both that the *protest paradigm* is used more conditionally than previously thought and that there have been overall changes in protest reporting in recent decades. There are limitations to the existing literature, however. The scope of studies has been rather narrow, focusing only on single, often radical, protest events or scrutinize the coverage surrounding a specific issue or movement. Furthermore, there are limitations to the theoretical foundation of the *protest paradigm*. Consequently, operationalisation of the paradigm and the way results have been interpreted differ substantially across studies, which has even led to contradicting findings regarding one protest event in the past.

The thesis uses a novel dataset of all articles published in eight national UK newspapers between 1992 and 2017 about domestic protests and demonstrations ($N = 27,496$). To analyse coverage in this large corpus, I use an innovative approach to framing analysis that combines best-practice manual coding techniques with supervised machine learning. Using this approach provides a strong methodological and theoretical foundation for the analysis of protest coverage: the operationalisation of frames is more explicit than in existing studies of the *protest paradigm* and frames are found inductively from the data, rather than being derived from decades old theory.

The analysis shows that a stable majority of articles uses frames linked to the *protest paradigm* throughout the time frame. At the same time, a substantial and growing number of articles employ legitimising frames — either on their own or co-existing with delegitimising framing. Specifically, I find seven distinct frames: four that follow the delegitimising patterns of the *protest paradigm*, two frames that legitimise protests and their message and one that is neutral. The results show that patterns of reporting about protest are not static and that the circumstances and features of protest events shape their coverage. Specifically, I find four main determinants for the use of the different media frames: (1.) violent protests get more delegitimising coverage, and less legitimising coverage; (2.) the goal of a protest matters for the kind of reporting it receives, yet relationships between frames and goals are complex and goals overall matter more for legitimising frames; (3.) protests receive less legitimising coverage from tabloid newspapers than from broadsheet outlets and one of the legitimising frames is used less often by right-wing media — which means that differences between outlet categories exist but are less pronounced than expected; and (4.) reports published more recently and longer after the start of an event have a higher chance of containing legitimising framing.

Overall, the thesis adds to existing knowledge on how the media frames protest over time and provides insights into the conditional logic with which journalists use different frames. Moreover, it develops a new approach to framing analysis combining manual and automated content analysis.

Table of Contents

Abstract **i**

List of Tables **vii**

List of Figures **x**

Acknowledgements **xii**

Author's Declaration **xiv**

1 Introduction **1**

2 Protest in the News **11**

2.1 Media Selection of Protest 12

2.2 Media Description of Protest — aka the *Protest Paradigm* 14

2.2.1 The *Protest Paradigm* 15

2.2.2 Empirical Knowledge 19

Chapter Conclusion 27

3 What Makes the Media Frame Protest? **29**

3.1 Media Framing 30

3.2 Random Reactions to Random Events? 35

3.3 Changing Media Ecology, Changing Protest Coverage? 38

3.4 What Drives Protest Coverage? 47

3.4.1 Event-Level Factors 49

3.4.2 Outlet-Level Factors 55

3.4.3 Time-Bound Factors 57

Chapter Conclusion 58

4 Research Design and Methods 61

4.1 Case Selection 62

4.2 Dependent Variable: Reporting of Protest 65

4.2.1 Definition of Protest 66

4.2.2 Measurement of Media Frames — Available Approaches 68

4.2.3 Semi-Automated Content Analysis of Media Frames — The New Approach 77

4.2.4 A Note on Sound Measurement 82

4.3 Independent Variables 86

4.3.1 Data 86

4.3.2 Matching Strategy 92

Chapter Conclusion 92

5 Building a Dataset of News on Protest Events 94

5.1 Newspaper Selection 95

5.2 A Note on Digital News Archives and *LexisNexis* 96

5.3 Pilot: Finding the Right Keywords 99

5.4 Collecting and Cleaning the Database 104

5.4.1 Removing Fragments of Data Entry 106

5.4.2 Removing Duplicated Entries 107

5.4.3 Removing Foreign Protest Events 108

5.4.4 Removing False Positives Using Topic Models 110

5.4.5 Removing Homonyms 112

5.4.6 Removing False Positives Using Machine Learning 113

Chapter Conclusion 115

6 Framing Protest **117**

6.1 Manual Coding 119

6.1.1 Frame Elements and Codebook 119

6.1.2 Coding Procedure 126

6.1.3 Inter-Coder Reliability 128

6.2 Frame Identification 131

6.2.1 Data Structure 133

6.2.2 Cluster Analysis 134

6.2.3 Factor Analysis 143

6.2.4 Interpretation of Frames 152

6.3 Frame Coding 160

6.3.1 Teaching the Computer to Code Frames 160

6.3.2 Results 166

6.4 Discussion and Summary 173

7 What Explains the Frames? 178

7.1 Data Processing and Models 179

7.2 Hypotheses Testing 187

7.2.1 Event-Level Factors 189

7.2.2 Outlet-Level Factors 195

7.2.3 Time-Bound Factors 197

7.3 Unpacking Frame Usage 198

7.3.1 Event-Level Factors 200

7.3.2 Outlet-Level Factors 207

7.3.3 Time-Bound Factors 208

7.4 Discussion and Summary 209

8 Conclusion 215

8.1 Findings 216

8.2 Contributions 220

8.3 Limitations and Future Avenues 223

References 227

Appendices 273

A Extended Codebook and Coding Instructions 274

B Additional Information on the Cleaning of the News Media Database . 302

C Regression Results for Change of Frame Share Over Time 310

D Recoded Protest Event Data 312

E Additional Models 342

List of Tables

2.1 Empirical Studies on the Protest Paradigm in Legacy Media, Ordered by Publication Year . . . 20

3.1 Hypotheses . . . 60

4.1 Confusion Matrix . . . 85

5.1 Selection of Newspaper Outlets for Data Collection . . . 96

5.2 Cleaning Steps . . . 105

5.3 Topic Examples . . . 111

5.4 Confusion Matrix for Validation of Naïve Bayes Model . . . 114

6.1 Codebook for Manual Content Analysis . . . 121

6.1 Codebook for Manual Content Analysis (continued) . . . 122

6.1 Codebook for Manual Content Analysis (continued) . . . 123

6.2 Inter-Coder Reliability for Identification Task . . . 130

6.3 Inter-Coder Reliability for Classification Task . . . 130

6.4 Removed and Merged Rare Codes . . . 131

6.5 Example Matrix of Coded Articles . . . 134

6.6 Confusion Matrix. k-Means vs. Manual Coding of Main Frame in an Article (Agreement in Green; Disagreement in Red) 142

6.7 Factor Loadings 146

6.8 Factor Loadings Full Model 150

6.9 Example Factor Scores 161

6.10 Example Document-Feature Matrix Including Class 162

6.11 Classification Performance of Frames (Top 3 Models) 165

7.1 Model Variables 180

7.2 Recoded Protester Demands 181

7.3 Distribution of Protester Demands 181

7.4 Comparison 186

7.5 Regression Results for Aggregated Frames 188

7.6 Regression Results for Individual Frames 201

B.1 Topic Model Used for Cleaning 302

B.2 Replaced Phrases 307

B.3 Phrases Used to Remove Articles 310

C.1 Regression Results for Change of Frame Share Over Time 311

D.1 Employed Subset of the MMP Data by Clark and Regan (2019) with Additional and Recoded Variables 312

E.1 Regression Results for Aggregated Frames (with all State Responses) . 343

E.2 Regression Results for Aggregated Frames (without The Times Articles) 344

E.3 Regression Results for Individual Frames (without The Times Articles) 345

E.4 Regression Results for Aggregated Frames (Newspaper Instead of Newspaper per Year as Level 2) 346

E.5 Regression Results for Individual Frames (Newspaper Instead of Newspaper per Year as Level 2) 347

E.6 Regression Results for Aggregated Frames (Robust Clustered Standard Errors) 348

E.7 Regression Results for Individual Frames (Robust Clustered Standard Errors) 349

List of Figures

3.1 Individuals using the Internet (percent of population), Selected Countries (Data Source: World Bank) 41

4.1 Frame Analysis Method 79

4.2 Number of Events in the PolDem Datasets over Time 90

4.3 Number of Protest Events in the UK per Year per Clark and Regan (2019) 90

5.1 Number of Times Each Keyword Was Relevant in the Pilot Sentences . 100

5.2 Variations of "protest" and "demonstrat". Larger Words Were Used More Often in the Pilot Articles 103

5.3 Number of Downloaded and Removed Articles per Year 105

6.1 Number of Times Each Code Was Assigned 132

6.2 Frequency Among All Indices for Best Cluster Solutions 136

6.3 Silhouette Plots for 2- and 3-Cluster-Solution 137

6.4 Heatmap of Cluster-Means for all Codes 138

6.5 Eigenvalues for Number of Factors and Percentage of Explained Variance 145

6.6 Eigenvalues for Number of Factors and Percentage of Explained Variance for the Full Model 149

6.7 Factor Loadings for Selected Frames 154

6.8 Number of Articles in Which Each Frame Is Present 167

6.9 Number of Articles in Which Any (De-)Legitimising Frame Is Present . 168

6.10 Percent of Total Articles per Year Containing a Frame 169

6.11 Percent of Total Articles per Year Containing a Frame, Significant Trends Highlighted 171

6.12 Percent of Total Articles per Year Containing (De-)Legitimising Coverage 172

7.1 Percent of Violent and Peaceful Protests Facing Different Kinds of State Responses 183

7.2 Average Marginal Effects for Significant Variables in Aggregated Frame Models 190

7.3 Presence of Frame in Percent of Articles About Protests with Different Goals 192

7.4 Predicted Probabilities of Using a Delegitimising Frame Before and After the Start of a Protest 199

7.5 Average Marginal Effects for Significant Variables in Individual Frame Models 202

7.6 Presence of Individual Frames in Percent of Articles About Protests with Different Goals 205

Acknowledgements

I am genuinely humbled by the support of the many people who encouraged and enabled me to start and finish this thesis. First and foremost, I thank Ana Langer, who convinced me to pursue a PhD, helped me to prepare my proposal, and became a mentor who guided me on becoming a proper academic (hopefully, she will see me succeed one day). I thank Brian Fogarty, who recommended me the first book on text analysis with R because he believed, for some unknown reason, that I could learn it in just a few months and use it in my MSc dissertation (which turned out to be true). I thank Sergi Pardos Prado, who joined my supervision team at a late stage but infused the project with a new level of enthusiasm and optimism when I thought that nothing really worked. This thesis is far better than it would have been without his guidance in my last year. I thank Philip Leifeld, from whom I learned a lot in a very short time about programming, R and academia while working as his Research Assistant. I am grateful for the reviewers of my annual reviews, attendees of my presentations at the university and at different academic conferences who provided valuable comments and questions that made me think about new aspects of my research. I thank the College of Social Sciences at the University of Glasgow for funding my research and making it possible to spend the years doing my PhD in Glasgow — which is still my favourite city in the world.

I thank Tim, who, for a brief time, was my office and flatmate and has since become a wonderful friend who always has an open ear for my worries. Also, I thank him and Ana Alonso for second-coding the material for my thesis and giving helpful hints on my codebook. I thank David Jofré for dragging me out of the office sometimes to blow off some steam and talk about the important things in life (like movies and hypothetical time travel concepts). I thank Louis Bujnoch, for among other things, hooking me to German bread after I had refused to appreciate it my entire life. I thank my third PhD colleague turned flatmate, Sacha, whose tendency to be happy is simply contagious. I thank Marina for looking at my regression models with me to make sure they are not complete nonsense. I thank Janina for keeping me company during our “digital writing retreats” via Zoom. I thank my friends and colleagues at the University of Glasgow: Alex, Anna, Blair, David, Dominic, Ekaterina, Evelyn, Gintare, Isabel, Luke, Stephen and Thuy, I miss you.

I thank especially my wonderful wife Alexandra, who encouraged me to seize the opportunity to study in Glasgow (twice) despite the challenges it put her through. We also shared the emotionally taxing task of writing a thesis, and there is no one I would rather have shared it with. Finally, I would like to express my sincere gratitude to my parents, Asta and Martin Gruber: Ich danke euch für eure Unterstützung und Ermutigung!

Chapter 1

Introduction

> "Like the tree falling unheard in the forest, there is no protest unless protest is perceived and projected" (Lipsky, 1968, p. 1151)

For ordinary citizens in modern representative democracies, the opportunities for participation in politics are limited. Citizens can contact their representative and try to convince them that an issue is worth their attention. Citizens can vote for a party to highlight their interest, assuming that one of the parties has identified the issue as something to work for. Citizens can join a party, rise through its ranks and — through skill and luck — get into a position where they themselves can speak in parliament or to the media. Or they might attempt to reach a level of fame or influence where their voices are more likely to be heard. Most commonly, though, public protest is the most immediate and lowest-threshold option available for ordinary citizens to participate in politics.

Public protest can thus be a powerful resource for formally relatively powerless groups to make their voices heard and influence public opinion and political agendas. From early protests by the anti-slavery and suffrage movements, to anti-poll tax, fox hunting demonstrations as well as anti-Iraq War and anti-Brexit protests, citizens marching in the streets are an important part of the history of Britain and other Western democracies. Protest can be a way for citizens to raise important issues and provide feedback to those in power. Protest can foster social change and contribute to the diversity of the marketplace of ideas. If protesters are heard and understood. In this capacity, protests are inherently important to democracy and, consequently, to political communication researchers. If democracies want to see citizens as more than people who can periodically cast a vote, they need to be included in the public debate. On the other hand, in modern mass democracies, not every voice can be heard by everyone all the time.

Protesters need to communicate to their audiences somehow. And indeed, there are several audiences for a protest, each of which needs to be targeted with a different goal in mind: participants receive messages that should motivate them and increase their participation while potential recruits and uninformed bystanders need to be convinced and converted to participants or supporters; opponents' messages should ideally be countered; the targets of a protest — such as government officials or corporate executives — and other institutions — such as religion or education — need to be convinced to instigate or promote change; and favourable coverage from the media needs to be secured (e.g., Lipsky, 1968; Tilly, 2012).

However, besides constituting an audience for protesters, the media can potentially also be a deciding factor for success or failure to reach other communication goals of a protest: as bystanders and participants of direct action usually make up an insignificant minority, the rest of the population learns of an event through mediatised messages. Mainstream media thus often establish the first contact between protesters and their audiences, even if direct messages, alternative or social media might become more important lines of communication once people know about a group or issue. Depending on the way and tone the media establish, this first contact can cause individuals to focus on certain aspects of an event rather than others (Druckman, 2001).

How protest is portrayed in the media is therefore a crucial issue for the communication potential of protesters and, more broadly, for the democratic process itself. Moreover, it is important for our understanding of the media. Do the media care for ideas and criticism by mostly non-elite protesters and offer them space on the "marketplace of ideas" (Gordon, 1997) or in the "public sphere" (Habermas, 1989)? Are they, in other words, a democratic counterweight to elites' power in public (media) discourse? Or are the media acting as an instrument of "social control" (Shoemaker, 1984) or institutions that carry out "system-supportive propaganda" (Herman and Chomsky, 1988)? Put differently, do they silence critical voices and protect the will of economic and political elites? The question of how the media treat protesters touches upon central questions about the media as a communication channel between different groups in modern democracies. The main aim of this thesis is therefore to add to existing knowledge about how protest is covered in mainstream news media.

Existing knowledge of the topic mainly focuses on two distinct struggles which protesters face — and most often lose — when trying to communicate via mainstream media (McCarthy et al., 1996; McLeod and Hertog, 1999; Smith et al., 2001): The first is a competition for media and audience *attention*. Despite the growing importance of social and alternative media, legacy media were — and are often considered to still be — the most important actors in selecting "newsworthy" events — a process referred to as gatekeeping (Friedrich et al., 2016; Russell Neuman et al., 2014). By doing so, news media outlets effectively decide what is publicly discussed and, therefore, if a larger audience beyond participants and bystanders will become aware of a protest's message (Weimann and Brosius, 2016) and potentially care about the issue — a process known as agenda setting (McCombs and Shaw, 1972). While important to keep in mind, and briefly discussed in Chapter 2, this thesis is *not* about which protests make it into the news. Rather, it is about the outcome of the second challenge for protesters: the struggle over the *intended meaning of their message*. Specifically, this thesis interrogates the way in which newspaper reporting framed protest events in the United Kingdom since 1992.

Doing so, the thesis contributes to existing knowledge in four main ways. First, it uses empirical data for the analysis that covers a broader scope and longer time frame than previous research. There already is a substantial body of literature that explores the coverage of protest in the media. This research is often subsumed under the term "protest paradigm" (Chan and Lee, 1984; McLeod and Hertog, 1992). The literature tends to argue that there is a tendency of journalists to portray protest events in a way so that protesters' messages appear as a side note, are completely omitted in favour of other details, such as the spectacle and conflict surrounding protests or the deviance of protesters, or are actively discredited by the media. However, empirical studies are primarily limited to specific protest movements or events (e.g. Iraq War or anarchist protests) and only cover brief periods. The lack of longitudinal studies is especially problematic since seminal work in the field is relatively old (e.g., Gitlin, 1980; Halloran et al., 1970; Murdock, 1973). Additionally, it has been argued that the last decades have brought significant changes to the media ecology and the way protests are organised and conducted — not least due to the revolutionary developments brought by the internet (Cottle, 2008). In order to capture differences between protests and developments over time, this thesis uses a new dataset of all articles that cover domestic protest and were published between 1992-2017 in one of eight major British newspapers.[1]

Second, I advance how protest coverage should be understood theoretically. The *protest paradigm* is often used as an ideal type to distinguish between coverage discussing ideas and reasons behind a protest event and coverage that delegitimises, marginalises or demonises protesters and their goals. However, the *protest paradigm* shows a theoretical gap, which hinders interpretation of empirical results: since it is based on the normative assumption that coverage following the paradigm will hinder a protest's chances to success, but does not spell out when this should be expected, it is left to the researcher's interpretation when to support the supposition of a *protest paradigm*. This explains, to some extent, why previous studies draw contradictory conclusions, some

[1] Specifically, *Daily Mail*, *Daily Telegraph*, *Financial Times*, *The Guardian/Observer*, *The Independent*, *The Mirror*, *The Sun* and *The Times* were used.

even while analysing the same material (see Chapter 2). To solve this problem, I introduce knowledge about framing and framing effects as the theoretical underpinning to evaluate and interpret media coverage of protest. Since it was found that framing has a limited effect on the audience if several competing interpretations of an event are available and accessible, I argue that the negative impact described in the *protest paradigm* literature should only be expected when delegitimising framing *dominates* protest coverage.

Third, I make a methodological contribution by introducing an innovative procedure to semi-automatically identify and code media frames in a large-scale media database. As the theoretical view on protest coverage called for a quantification of reporting patterns, the challenge arose how to code the large dataset in a reliable and valid way. Furthermore, while the *protest paradigm* literature offers a number of media frames, how they were identified is often not entirely clear and frames might be specific to the circumstances of protests that scholars studied at the time. I combine best practice methods of manually *identifying* frames inductively with natural language processing and machine learning approaches to *code* frames in the large dataset of media reports covered in this thesis.

Fourth, using the outcome of the framing analysis, this thesis analyses which kind of coverage different protests receive by different media outlets and over time. Specifically, the literature suggests a number of factors that are thought to "trigger" (Lee, 2014) coverage following the *protest paradigm* on several levels: *Protest Goal*, *Protester Violence*, *State Repression*, *Newspaper Ideology*, *Newspaper Type*, *Ideological Divide*, *Days Since Start*, and *Year of Protest*. The final contribution of this thesis is thus to identify and systematically *test* which — if any — factors explain which of the identified frames are used and to which degree. While some of these factors have been assessed before, this thesis is, to my knowledge, the first study to compare all of these outlet specific, time specific, and protest specific variables.

Overall, three research questions guide this thesis:

RQ1: How do British newspapers frame the coverage of domestic protest events?

RQ2: How — if at all — did the framing of protest reporting change over the last 26 years?

RQ3: Which factors explain the choice of frames by the news media when covering domestic protest events?

The first question aims to provide a comprehensive set of frames and test how prevalent they are in reporting. The second question aims to shed light on fluctuations or substantial change in the distribution of frames about protest over time. The third research question aims to determine which frames are associated with any of the abovementioned variables of a protest event, the outlet covering it and the point in time protest coverage is published.

To address these research questions, I chose a longitudinal case study design of a single country. While a cross-country comparison would provide the opportunity to test the influence of the differences between media systems on protest coverage (Hallin and Mancini, 2004), the design used here increases internal validity (Pepinsky, 2019) and allows for greater nuance in the analysis of the different treatment of events and change over time. The longitudinal design, conversely, allows investigating the dynamics brought by apparent changes in the media and society of the UK. By choosing the United Kingdom as a case, it is possible to link the results to seminal work about protest reporting from the 1970s and 80s and compare them to the numerous studies from the US — as cross-country research already exists (Dardis, 2006b). Moreover, since the UK has an especially partisan media landscape (e.g., Kuhn, 2007), it enables me to draw conclusions about the influence of the ideology and type of an outlet on the framing of a story about protest.

Protest is understood here relatively broadly, as *a collective overt public expression that either articulates grievances against or support for one or multiple targets (i.e.,* *policies, institutions or behaviours) in order to either directly influence an institution's decisionmakers or the knowledge, attitudes, and behaviours of the public* (see Chapter 4 for details).

As a result of the broad definition and the long time frame, the novel dataset created for the thesis contains more than 27,000 newspaper articles about protest. This made a purely manual content analysis unfeasible. However, as mentioned above, the procedure which was used to determine the framing of articles in the dataset is semi-automatic. Specifically, this means that after a manual phase of finding individual frame elements and coding them manually in a sample of the data, dimension reduction techniques are used to determine the number and attributes of frames. This results in high reliability and validity of the identified frames as subjective human judgement is significantly reduced compared to methods where humans code frames directly and holistically in text (Matthes and Kohring, 2008). The empirically induced frames are then coded automatically in the articles which were not included in the sample yet, using natural language processing in combination with machine learning.

The analysis shows that a stable majority of articles uses frames linked to the *protest paradigm* throughout the time frame. At the same time, a substantial and growing number of articles employ legitimising frames. Specifically, I find seven distinct frames: four that follow the delegitimising patterns of the *protest paradigm*, two frames that legitimise protests and their message and one that is neutral. The results show that patterns of reporting about protest are not static and that the circumstances and features of protest events shape their coverage. In particular, I find that they are conditioned by event- and outlet-level as well as time-bound variables: violent protests get more delegitimising coverage, and less legitimising coverage; the goal of a protest matters for the kind of reporting it receives, yet relationships between frames and goals are complex and goals overall matter more for legitimising frames; protests receive less legitimising coverage from tabloid newspapers than from broadsheet outlets and one of the legitimising frames is used less often by right-wing media — which means that differences between outlet categories exist but are less pronounced than expected; and reports published more recently and longer after the start of an event have a higher chance of containing legitimising framing.

The remaining thesis is organised into seven chapters. Chapter 2 will identify the central debates and gaps in the literature. As hinted above, the literature on news coverage of protest describes two separate potential problems: a media selection bias, meaning that the media prefer to cover some protests over others, and a media description bias, meaning that the media highlight some aspects of a protest over others. The most important concept which aims to make sense of the latter problem is the *protest paradigm*. It encompasses much of the theoretical work and informs nearly all of the empirical work done in the field in the last three decades. Neither the theoretical nor the empirical aspects of research in this area is without limitations, however.

Chapter 3 provides a theoretical underpinning to study protest coverage and introduces external factors that are thought to influence how the media frames protest events. I argue that theoretical and empirical advances in *framing* research offer a tangible way to systematically analyse media coverage and interpret the findings. I also examine the changes that occurred on the media landscape, how protests unfold and how the broader society judges protesters. Taken together, I argue that these changes cast doubt on previous conclusions about protest coverage. I then introduce the independent variables of this thesis, which represent factors thought to condition coverage of protest on multiple levels: event characteristics, features of the reporting outlet and change of reporting over time.

Chapter 4 outlines the research design and methods and provides information about the database from which the independent variables are drawn. The chapter discusses why I chose newspaper coverage about protest in the UK from 1992 to the end of 2017 as the case for this thesis and what I understand as *protest* in this study. The analysis strategy for this thesis is a two-step procedure: the first step measures the framing of each article in the dataset of news coverage of protest. To do that, I develop a new method of framing analysis based on the best practice method of *frame identification* and a procedure of *frame coding* based on Automated Content Analysis (ACA). Central to the *frame identification* step is the understanding of frames as latent dimensions developed here. Based on this, frames can be found by coding frame elements and using dimension reduction techniques to uncover them. In the *frame coding* step, I use 500 randomly selected and manually coded articles to train machine learning models that subsequently code the remaining data. The second step is to test which of the included independent variables determine the use of the identified frames in news media coverage of protest.

Chapter 5 describes how the comprehensive dataset of newspaper reports used in this thesis was compiled. I provide details about the choices made regarding the digital newspaper archive, newspaper outlets and how I used a pilot study to empirically determine an optimal set of keywords to query data. When constructing the database, the goal was that it should contain all articles about protest published in the eight selected national UK newspapers during the employed time frame. However, since the dataset initially included many irrelevant articles, I used different methods to clean the dataset. Ultimately, I removed 95% of the initially downloaded articles as they did not contain coverage of domestic protest as defined in this thesis.

Chapter 6 describes the first step in the two-step research design employed in this thesis: the main frames used by UK media to portray protest are first *identified* and then *coded* in the news media dataset. I will explain how I developed a codebook, coded a random set of 500 articles manually and established intercoder reliability. Based on this data, the frame *identification* step was performed using first cluster and then factor analysis. I then reused the factor scores from factor analysis in the frame *coding* step. The information about the presence or absence of each frame in the manually coded sample was then used to train different machine learning models and use the best ones to code frames in the remaining newspaper articles. The chapter then answers **RQ 1** and **RQ 2**.

Chapter 7 represents the second step in the two-step research design: the measured framing in the coded articles is explained through the independent variables. Specifically, I use multilevel logistic regression, with newspaper-years as the second level and one model for each frame, to test the relationships between the selected independent variables and the presence of frames in the articles. This answers **RQ 3**. Besides the hypotheses developed in Chapter 3, I test several theories on news coverage of protest and discuss how they fare against my results.

The final chapter summarises and discusses key findings, draws implications and discusses limitations.

Chapter 2

Protest in the News

For most of us, the media are the primary source of knowledge about the world and what is happening in other societies and states as well as our own (e.g., Luhmann, 2000). In effect, the crowd of spectators present at a demonstration has lost most of its importance compared to the many times larger audiences perceiving the event through the lens of the mass media (Tilly, 1995). That means that the decision about failure or success of a protest — that is if it is able to inspire change or at least spark a public discourse about a topic (Amenta et al., 2017) — is mainly decided in indirect mediated encounters "among contenders in the arena of the mass media public sphere" (Koopmans, 2004, p. 367). The media do this as they help shape the public's attitudes towards new issues and often also provide the starting point from which these attitudes are formed. More specifically, the media are commonly attributed two crucial powers: the power to influence what people are talking about, known as *agenda setting*, and the means and narratives in which people talk about it, often described as *framing* (i.e., the description of events).[2]

[2] A third effect or power of the media often mentioned is *priming*, which is thought to influence what types of considerations (e.g., economic) the audience uses when thinking about a particular issue put on the agenda (e.g., Iyengar and Kinder, 1987). Since the focus of this thesis lies on media content instead of media effects, I omit *priming* in this thesis.

This chapter presents what we know about *agenda setting* and *framing* patterns in protest coverage. The first Section 2.1 covers the selection and salience of protest events, while Section 2.2 focuses on the description of protest in the media. Generally speaking, this literature confirms the broader picture that the media arena heavily privileges those with economic or political power or other forms of authority while disadvantaging ordinary citizens (e.g., Bennett, 1990; Entman, 2004; Wolfsfeld, 1997). However, this finding is not unconditional, and recent research suggests that the media arena is not as impenetrable for protesters as has been described in the past.

2.1 Media Selection of Protest

One of the principal powers the media are attributed is *agenda setting*. The idea behind *agenda setting* is that the more the media cover certain stories and issues, the more likely people are to talk and think about them: “The press may not be successful much of the time in telling people what to think, but it is stunningly successful in telling its readers what to think about” (McCombs and Shaw, 1972, p. 177). This often influences politicians as well, since political actors see media attention as a proxy for public opinion and use high media attention on a subject to manoeuvre the topics they “own” into the spotlight (Elmelund-Præstekær and Wien, 2008). However, setting the agenda also implies that news media can determine what people do not think about. As Lipsky (1968) put, “there is no protest unless protest is perceived and projected” (p. 1151). In other words, if an issue is not covered at all by the media, the public has no chance to hear about it — unless people happen to be bystanders at an actual event. By including some issues while missing out on others, the news media are thus quite powerful in deciding what is — and what is not — discussed in political discourse. In other words, media are the gatekeepers of public discourse (e.g., Livingston and Bennett, 2003). The absence of scrutiny by the public, though, can entice officials to ignore a topic or follow the insinuation of organised groups unchecked (Burstein, 2006). Legacy media were and often still are particularly important for agenda setting, as they provide a main source for investigative journalism and are able to amplify or sustain public pressure, even though alternative and social media have gained importance over the years (Langer and Gruber, 2021; Shoemaker and Reese, 2014, pp. 40-42; Curran et al., 2013). This also explains why protesters still actively seek the attention of legacy media despite the undeniable decline in audiences (Chadwick and Dennis, 2017).

Consequently, there is a substantial body of literature investigating how media *select* and give salience to some events over others. One way to do so is to compare official resources that keep track of protest events (such as permits and police records) with media reports to see how many and which events make it to the news. These studies agree that the large majority of demonstrations do not receive any mainstream coverage whatsoever (Boyle et al., 2012; Hocke, 1999; McCarthy et al., 1996; Oliver and Maney, 2000; Oliver and Meyer, 1999; Wouters, 2015b). They furthermore suggest that the media are guilty of a selection bias, which means that not all events have an equal opportunity to be covered as certain events are given precedence over others. This bias, however, is then reproduced by protest coverage studies (McCarthy et al., 1996) — including this thesis, which focuses on how the media treat protests once they made it into the news. So why are some events covered less likely or not at all? The basic argument is, unsurprisingly, that some events are seen as more newsworthy by the press and thus have a higher chance to be covered in more detail. It is also noteworthy that protests do not simply struggle against an unwillingness of reporters to cover them. They compete against other stories — maybe even other protests — for a finite amount of airtime, column inches or audience attention (Gavin, 2010).

Specifically, McCarthy et al. (1996) find that several factors predict the likeliness to which an event will be covered: the size of a demonstration and how prominent the issue of a protest already is in the current media discussion (also see Hocke, 1999; Hug and Wisler, 1998). Oliver and Meyer (1999) confirm this assessment and added that events that involve conflict — either with the police or counter-demonstrators — are sponsored by a business group, or occurred in a central location have a higher chance to make it into the news (also see Oliver and Maney, 2000). Mueller (1997) adds that in the international context, geographical distance from a newspaper's headquarters to a protest determines how likely it is to be covered. Barranco and Wisler (1999) then show that proximity has the same effect in a national context.

Reviewing the body of research, Earl et al. (2004) offer a comprehensive typology. The selection of protest events as news stories is determined by: *event characteristics*, especially size of a protest, if a protest was supported by a sponsor, such as business groups, NGOs or celebrities, and whether violence or other disruptive events occurred; *news outlet characteristics*, such as proximity of the outlet's headquarters to an event; and *protest issue characteristics*, especially if the issue was already on the media agenda. More recent reviews come to similar conclusions (e.g., Amenta et al., 2017; Ortiz et al., 2005; Wouters, 2015b).

As mentioned above, the focus of this thesis is on the content, not inclusion or salience of coverage on specific protests. However, the fact that most protest does not make it into the news and which features make it more likely to receive coverage does help to understand some of the choices protesters make. Why do some protests employ disruptive and militant tactics, for example, even though this can lead to delegitimising coverage, as explained below? The subsequent coverage might be negative, but disruption and drama can put a protest on the map (Boyle et al., 2012; Weaver and Scacco, 2012). Noteworthy is also that the typology of variables offered by Earl et al. (2004) informed the typology I present in Section 3.4, which is, hence, similar. Event and news agency characteristics do, as I show later, have an influence not just on the selection of protest events but also condition the kind of coverage protests receive.

2.2 Media Description of Protest — aka the *Protest Paradigm*

> "In newsmaking, journalists do not merely use culturally determined definitions, they also have to fit new situations into old definitions. It is in their power to place people and events into the existing categories of hero, villain, good and bad, and thus to invest their stories with the authority of mythological truth" (Bird and Dardenne, 1988, p. 80).

When it comes to the account of protest in the news, the critique of a biased mainstream media landscape is widespread among activist circles (e.g., McCurdy, 2008; Rucht, 2013). In academia, this notion is often repeated and supported by a wide range of studies. Over the years, researchers found that when covering protest events, the media often follow certain typical patterns that amount to the so-called "journalistic protest paradigm" (McLeod and Hertog, 1992). The concept seems to be employed, to some degree, by most media scholars analysing the characteristics of protest coverage. The following section describes the *protest paradigm* as one of the central theoretical concepts in studies about media coverage of protest before reviewing empirical results from the field and identifying existing gaps and shortcomings of previous research.

2.2.1 The *Protest Paradigm*

After the salience of protest coverage increased following the civil rights era, landmark works of Halloran et al. (1970) and Murdock (1973) in the UK and Gitlin (1980) in the US, showed that protests in the 1960s had often been delegitimised in reports by mainstream media. Shortly after Gitlin's study, Chan and Lee (1984) comprised the theoretical explanations and empirical observations into a single concept. Relying loosely on the "paradigm" concept by Kuhn (1970), Chan and Lee (1984) suggested that a " 'metaphysical' world view or a gestalt" (p. 187) shapes what journalists define as entities of concern, indicate "to journalists where to look (and where not to look), and informs them about what to discover" (p. 187). This view, assumed to be held by the bulk of professional journalists, leads reporting to fall in line with pre-defined, ideological informed patterns of reporting whenever reporters make sense of a story. The concept was taken up by McLeod and Hertog (1992), who coined the expression "protest paradigm", which was defined as "routinized journalistic paradigm for covering social protest" (p. 206).

However, what exactly comprises the *protest paradigm*? In its most advanced, and probably broadest, conceptualisation, McLeod and Hertog (1999) state that the *protest paradigm* combines *unfavourable story framing*, *reliance on official sources and official definitions* and the *invocation of public opinion* (e.g., the protesters are a minority). Specifically, McLeod and Hertog (1999) list a number of common narratives which are repeated in protest coverage: the *violent crime*, *property crime* and *riot* narratives focus on the violence committed by protesters against people and property and imply a general lawlessness during events, often accompanied by clashes with the police; *carnival*, *freak show*, *romper room* and *moral decay* highlight the spectacle of protest and the oddity, deviance and naïvety of protesters; and *storm watch* warns the audience about possible — often widely overstated — threats of protests for bystanders and the society at large.

The *protest paradigm* is closely linked to broader theories of newsmaking, which suggest that reporters tend to place new events into an internal "mental catalogue of news story themes, including how the 'plot' will actually unravel and who the key actors are likely to be" (Berkowitz, 1992, p. 83), which helps them to streamline the news gathering and decision-making processes. Narratives and procedures which have worked well in the past therefore influence how new stories are covered. Therefore, celebrities, ordinary people and victims often receive standardised roles within the narrative of a news story (Langer, 1998). Dayan and Katz (1992) even suggest that just three forms of scripted narratives can characterise the whole spectrum of news reporting *Contests or epic* *contests* (most often in sports and politics), *Conquests* (drastic one-time events that overcome the rules) and *Coronations* (ceremonial events such as actual coronations or the inauguration of the president). While these highly abstract categories might overstate the point, it seems clear that cultural settings and organisational routines contribute to creating a generalising "media logic" which tends to standardise the form, features, and frames of news stories (Altheide, 1995).

Individual journalists, as well as media organisations, are therefore eagerly "routinizing the unexpected" (Tuchman, 1973, p. 110) by creating practical and effective procedural rules, routines and mental catalogues of news story themes. This is deemed an effect of necessity, as journalistic work is generally characterised by tight deadlines, limited resources, a basically infinite supply of raw information, a high uncertainty about what is important and interesting enough to be news and the danger of possible libel suits or scolding by superiors, let alone considerations of what the audience might want. Essentially then, the *protest paradigm* is one of these mental catalogues, just specifically for protest. However, what is thought to be worrisome about the *protest paradigm* is that while the existence of alternative narratives is not strictly denied, and McLeod and Hertog (1999) postulate the existence of mixed, balanced and sympathetic reporting, this coverage is supposed to be rare in mainstream outlets.

The basic idea of the *protest paradigm* is, therefore, that while there is assumed to be diverse groups behind, and many different objectives of protests, general influences on media production are working against protesters as pre-defined routines lead to coverage which diminishes the chances of a protest's success. This pattern is also highlighted by the *reliance on official sources and official definitions* when it comes to crafting reports about protest events (e.g., Dardis, 2006a; McLeod and Hertog, 1992, 1999). As before, this is related to broader trends identified by research on the sociology of newsmaking: officials are often the most frequently used sources, are regarded as experts with access to high-quality or first-hand information and are generally regarded as newsworthy by default (e.g., Bennett, 1990). Furthermore, officials might be included in a story to establish a certain balance to protesters' message, which usually challenges a specific policy. However, since social protests usually try to change present circumstances, the selection of official's interpretations — who usually want to maintain the status quo — undermines their efforts.

Similarly, *invoking public opinion* is described as a key feature of the *protest paradigm* but is also a standard tool in reporting in general. Nevertheless, it might work systematically against protesters when, instead of opinion polls — which are often unavailable — reporters use bystanders as an implied proxy for public opinion. As McLeod and Hertog (1999) point out: "Almost by definition, the onlooking bystanders are at best indifferent to the protest and at times hostile" (p. 317) as they would otherwise join in on the protest or would be recognised as a participant by a reporter (also McLeod and Hertog, 1992). Reporters might do this simply in anticipation of what their audience wants to read or how their opinion on the issue might be. Additionally, editorial decisions might demand a report to be shorter, more graphic or feature a newsworthy "peg", thus shifting the action into focus and away from the protest's message (Smith et al., 2001; McCarthy et al., 1996; Ryan, 1991; also see Shoemaker and Reese, 2014).

Notable, journalistic routines are only one of several common explanations for the existence of a *protest paradigm* in reporting. Smith et al. (2001) suggest that several authors mention ideology and power structures to explain why dissenting voices like protesters receive the coverage they receive. In short, the ideological explanations state that the newsmaking process is organised in terms of the cultural reproduction of power structures. In this view, economic and political elites subtly control public understandings of issues to stabilise the status quo. Since groups who organise protests usually dissent with those in power, elites would use their influence to make sure journalists continue to employ the *protest paradigm* (e.g., Boykoff, 2007; Chan and Lee, 1984; Weaver and Scacco, 2012).

Structural explanations take this one step further by saying that the media "carry out a system-supportive propaganda function by reliance on market forces, internalised assumptions, and self-censorship, and without significant overt coercion" (Herman and Chomsky, 1988, p. 306). This explanation supposes that protests do not receive supportive coverage or any coverage at all if this would hurt advertising opportunities or other valuable connections to those with economic or political power. Therefore, market interests of news companies crowd out ideas threatening the status quo (e.g., Gamson et al., 1992; Herman, 1995). As I will discuss in Section 3.4, I employ several variables based on ideology and power structures to test why certain protests receive certain coverage.

How do the practices described above affect protests? McLeod and Hertog (1999) suggest that regular employment of the *protest paradigm* can contribute to three processes: delegitimisation, marginalisation, and demonisation of protest events, protest movements, or protest as a political resource. In short, delegitimisation means that news coverage undermines the legitimacy of a protest group or issues; marginalisation means that coverage portrays a protest as smaller and more deviant so that citizens overlook that their own concerns match those of the protesters; demonisation means that the potential threat of a protest is exaggerated so that protesters appear as dangerous and a threat to public safety, scaring away potential supporters. In other words, as groups need to establish themselves as a legitimate voice in the political discussion to initiate change, the above-mentioned reporting patterns can compromise the chance of a group to establish their message in the public debate (Arpan et al., 2006; Detenber et al., 2007; McLeod, 1995). However, is this what actually happens? The next section reviews existing studies to establish what we already know about the existence and prevalence of the *protest paradigm* in reporting.

2.2.2 Empirical Knowledge

So what has been found in practice about the characteristics of the coverage of protest? As mentioned above, there is a close overlap between the literature on the coverage of protest in the mainstream media and the literature on the *protest paradigm*. Furthermore, even the studies which depart from this tradition follow much of the same logic when analysing media narratives about protests. Rosie and Gorringe (2009), for example, compare the messages of a large-scale protest in Edinburgh during the G8 Summit in 2005 with the media frames covering it. They found a strong tendency of the media to cover just the spectacle or fall into a delegitimising, “pre-existing frame for G8-related protest environments” (p. 51), which fits the descriptions of the paradigm approach. Boykoff (2006), as a second example, assessed the most dominant frames employed during the World Trade Organization protests in Seattle in 1999. He found that a *Violence* frame, a *Disruption* frame, a *Freak* frame, and *Ignorance* frame and what he calls an “*Amalgam of Grievances*” frame defined the coverage. The first four of these frames fit closely with what the *protest paradigm* already describes, while the “Amalgam of Grievances Frame” basically accused the protests of having no clear message — a variation of the “Romper Room” theme. The studies by Smith et al. (2001), Jha (2007) and Elmasry and el Nawawy (2017) form another category: They basically employ the theoretical framework of the *protest paradigm* — including the literature cited in the previous section — find support for it, but do not refer to the concept directly.

Table 2.1: Empirical Studies on the Protest Paradigm in Legacy Media, Ordered by Publication Year

Study	Country	Case Period	Support for PP	Measured
McLeod and Hertog (1992)	US	1986–1987	Yes	Other
Smith et al. (2001)	US	1982, 1991	Yes	Episodic vs. Thematic
McFarlane and Hay (2003)	Australia	1999	Yes	Marginalisation Devices
Boyle et al. (2004)	US	1960–1999	Partly	Episodic vs. Thematic
Boyle et al. (2005)	US	1960–1999	Yes	Valence
Brasted (2005)	US	1968	Yes	Marginalisation Devices
Dardis (2006a)	US	2002–2003	Partly	Marginalisation Devices, Valence

Table 2.1: Empirical Studies on the Protest Paradigm in Legacy Media, Ordered by Publication Year *(continued)*

Study	Country	Case Period	Support for PP	Measured
Dardis (2006b)	UK, US	2002–2003	Partly	Marginalisation Devices, Valence
Gavin (2007)	UK	2000	Yes	Marginalisation Devices
Jha (2007)	US	1967–1999	Yes	Episodic vs. Thematic
McLeod (2007)	US	2006	No	Marginalisation Devices
Boyle and Armstrong (2009)	US	1960–2006	Yes	Valence
Di Cicco (2010)	US	1967–2007	Yes	Other
Gavin (2010)	UK	2008	Yes	Marginalisation Devices
Edgerly et al. (2011)	US	2006	Yes	Marginalisation Devices
Harlow and Johnson (2011)	US, Egypt	2011	Yes	Marginalisation Devices
Sela-Shayovitz and Hasisi (2011)	Israel	2000, 2007	Yes	Other
Boyle et al. (2012)	International	2007–2009	Yes	Valence
Corrigall-Brown and Wilkes (2012)	Canada	1990	No	Other
Weaver and Scacco (2012)	US	2009–2010	Yes	Marginalisation Devices
Xu (2013)	US	2011–2012	Yes	Marginalisation Devices
Young (2013)	US	2011–2012	Yes	Marginalisation Devices
Baysha (2014)	US, Russia	2011	Yes	Media Frames
Lee (2014)	Hong Kong	2001–2012	Yes	Marginalisation Devices
Ghobrial and Wilkins (2015)	Egypt, Saudi Arabia, Tunisia, US	2011	Yes	Other
Gottlieb (2015)	US	2011–2014	Partly	Media Frames
Hughes and Mellado (2015)	Chile	1990–2011	Partly	Marginalisation Devices
Spyridou (2015)	Cyprus	2013	Yes	Media Frames

Table 2.1: Empirical Studies on the Protest Paradigm in Legacy Media, Ordered by Publication Year *(continued)*

Study	Country	Case Period	Support for PP	Measured
Wouters (2015a)	Belgium	2003–2010	Partly	Other
Wouters (2015b)	Belgium	2003–2010	Partly	Episodic vs. Thematic
Coulter et al. (2016)	Ireland	2002–2003	Yes	Valence
Oz (2016)	Turkey	2013	Yes	Media Frames
Power et al. (2016)	Ireland	2014–2015	Yes	Marginalisation Devices
Reul et al. (2016)	Belgium	2011	Yes	Marginalisation Devices
Shahin et al. (2016)	Brazil, China, India	2011–2013	Partly	Marginalisation Devices
Veneti et al. (2016)	Hong Kong	2014–2015	Partly	Media Frames
Elmasry and el Nawawy (2017)	US	2014–2015	No	Other
Kyriakidou et al. (2017)	Spain, Greece, Germany	2011	No	Marginalisation Devices
Kyriakidou and Olivas Osuna (2017)	Spain, Greece	2011	No	Marginalisation Devices
Leopold and Bell (2017)	US	2014	Yes	Marginalisation Devices
Trivundža and Brlek (2017)	Slovenia	2012–2013	Yes	Marginalisation Devices
Ismail et al. (2019)	US	2014, 2017	Partly	Other
Kilgo et al. (2019)	US	2013, 2014	Yes	Marginalisation Devices, Media Frames
Mourão (2019)	Brazil	2013	No	Marginalisation Devices
Coombs et al. (2020)	US	2016	Yes	Media Frames
De Cillia and McCurdy (2020)	Canada	2018–2019	No	Marginalisation Devices
Gil-Lopez (2020)	US	1998–2017	Yes	Marginalisation Devices
Harlow (2020)	US	2017	Yes	Media Frames
Harlow et al. (2020)	World-wide	2014	Partly	Framing, Sourcing, Marginalizing Devices
Kim and Shahin (2020)	Korea US	2016–2017	Partly	Media Frames

Table 2.1: Empirical Studies on the Protest Paradigm in Legacy Media, Ordered by Publication Year *(continued)*

Study	Country	Case Period	Support for PP	Measured
Papaioannou (2020)	Cyprus	2014	No	Media Frames
Umamaheswar (2020)	US	2016	Yes	Marginalisation Devices

Among studies that look for the *protest paradigm*, there seems to be a general agreement that coverage of the studied events contains at least some patterns described by the *protest paradigm* literature. In line with McLeod and Hertog (1999), most of these studies present the *protest paradigm* as the default condition for reporting, yet do not indicate that mainstream news media *always* describe protest in such a way. When it comes to how strictly media follow the *protest paradigm* across different protest groups, in different situations and at different points in time, however, there is less agreement in the existing literature. Table 2.1 shows a systematic examination of 52 studies which worked with the paradigm since 1984 — the year in which Chan and Lee coined the term. The assumption that legacy media heavily employ the devices from the *protest paradigm* in their coverage is largely confirmed: Table 2.1 shows that 32 studies concluded that the coverage they examined was in line with the paradigm. However, 12 found only partial support and eight concluded they found contradicting evidence.

In part, this disagreement is due to a lack of consensus on how the *protest paradigm* should be conceptualised and measured. Table 2.1 shows that there are at least three different, although closely related, patterns scrutinised in the literature. First, 26 of the studies in Table 2.1 measure the use of *marginalisation devices*, which means the unfavourable features of coverage described above. Second, studies examine if coverage focuses on the features and methods of a protest rather than the issue protesters want to raise attention for, often described as episodic in contrast to thematic coverage.

Third, six studies measure the valence of articles, meaning if reports are critical or supportive of a protest.

This lack of clear measures and conceptualisation makes the studies hard to compare directly and, unfortunately, might explain some of the differing conclusions regarding the prevalence of the *protest paradigm*. Two of the studies claiming partial evidence, namely Wouters (2015b) and Boyle et al. (2004), analyse thematic versus episodic coverage, which none of the supporting studies do. The result of one of the two no-studies also seems closely connected to how they conceptualise the paradigm: Corrigall-Brown and Wilkes (2012) assessed pictures of the 1990 "Oka Crisis" and found that protesters are portrayed more prominently and also to be in charge of a situation more often than officials. However, "prominence" is a concept no other study employs. In contrast, most studies that find support for the *protest paradigm* explicitly test the occurrence of marginalisation devices. Furthermore, several of them appear to take the fact that they find any stories that employ these devices as evidence for the paradigm's importance (Boyle et al., 2005; Brasted, 2005; Harlow and Johnson, 2011; McFarlane and Hay, 2003; McLeod and Hertog, 1992; Weaver and Scacco, 2012).

The conflicting conclusions can be attributed, at least partly, to a theoretical gap in the *protest paradigm*: from a normative view the *protest paradigm* is seen as problematic since it is assumed to lead to the above-mentioned delegitimisation, marginalisation, and demonisation of protest. However, should this be expected as soon as *any* coverage following the paradigm is found? In other words, what is the "cut-off point" at which studies should claim support for the existence of the *protest paradigm*? Is it bad if a majority of stories include elements of the paradigm — that is, at least 51% of the articles? Or is *protest paradigm*-coverage only diminishing protests if it dominates the coverage unchallenged by supportive arguments? The theory does not directly speak to this issue, leaving room for different interpretations of results. The comparison of the outcome of two studies helps to exemplify this issue: McLeod (2007) and Edgerly et al. (2011) incidentally, and seemingly unaware of each other, chose the same case to examine *protest paradigm* coverage — the “Day Without Immigrants” on May 1, 2006 in Los Angeles. Yet, the studies provide contradictory conclusions: While McLeod (2007) asserts that expectations derived from the *protest paradigm* were *not* met in the news media coverage, Edgerly et al. (2011) describe a different finding:

> “the protest paradigm continued to be a powerful organizing principle in media coverage of the protests. Most significantly, organizers did not generate comprehensive coverage of their legislative goals in the mainstream press, or overcome the episodic and tactical framing of most reporting on political protest” (p. 329).

While McLeod (2007) confirms the lack of “detailed explanation of the issues behind the protest” (p. 190), he seems to find it less important and not to be evidence for the *protest paradigm*.

This seems key as many studies do not even report how prevalent the paradigm actually was in their coverage. As Dardis (2006a) points out, his two studies from 2006 are the first to quantify the relative prominence of marginalisation devices. He found that stories assessed as being negative featured 3.33 marginalisation devices per story while positive coverage still contained 2.02 devices per story. In fact, he asserts that only four of fourteen tested marginalisation devices were disproportionately associated with overall negative coverage. Hence, the sheer presence of the devices described above does not automatically determine that a protest is marginalised by the media. This was confirmed recently by Kyriakidou and Olivas Osuna (2017), who found that while most coverage of the Indignados movement’s protests employed a spectacle frame, the tone of the reporting was positive, and the celebratory element of the demonstrations helped to create a supportive image of the movement in the press. Yet, what earlier studies often did was to assess if *any* devices from the paradigm were present in the articles. Yet, as Kyriakidou and Olivas Osuna (2017) have shown, the mere presence of devices from the paradigm in protest coverage might not automatically lead to marginalisation, delegitimisation or demonisation of a protest event.

Besides these issues, how generalisable are previous findings? Most of the studies examine narrow case studies of single, often radical, protest events or scrutinise the coverage surrounding a specific issue or movement. This is, of course, fine when exploring if certain features of a protest lead to a particular coverage. McLeod (2007), one of the two studies that found no evidence for the paradigm, specifically selected a large, mainstream movement lacking the features often thought to amplify the usage of the paradigm to test if he would still find marginalising coverage. However, as there is no ceteris paribus comparison, it is hard to assess if McLeod's (2007) findings about an event in 2006 are due to the specific character of the protest event or if other factors, for example, in newsmaking or event organisation have changed since his work in the 1990s (esp., McLeod and Detenber, 1999; McLeod and Hertog, 1992). For other studies, the restriction to specific cases also limits the scope of their findings. Dardis' (2006a; 2006b) studies, for example, are specific to anti-war or even just anti-Iraq-War protests. It is thus questionable if they can serve as suitable evidence in the more general debate about the persistence of the paradigm in mainstream news coverage — which Dardis also makes explicit in his conclusion.

Another set of studies considers a broader scope of protests but is still limited in other ways. Boyle et al. (2004), for example, consider a wide variety of protest events and also take a long period of time (1960-1999) into consideration. However, they limit their research to five local Wisconsin newspapers that might be argued to cover stories different from national newspapers in the US. Another example is the study by Shahin et al. (2016), who, commendably, also add a comparative element to their study by examining protest coverage in three non-Western countries to add to the predominantly US-focused body of knowledge. However, they selected 30-day periods for their sampling and do not take into account how coverage changes over months or years, for example, due to journalistic issue-attention cycles or developments in the wider media ecology and society. The lack of longitudinal studies appears as a shortcoming as the media ecology is thought to have undergone some considerable change in the last two decades (see Section 3.3).

Most sampling strategies are also relatively narrow: in an otherwise well-designed study, Boyle et al. (2012), for example, retrieve articles "by a Lexis-Nexis search using the key word *protester* to search headlines" (p. 132). Yet, they do not further explain the step of limiting the scope to "protester" instead of "protest" or why they search just in headlines. Worryingly, this might be due to a resource problem rather than a deliberate decision: the sample sizes in the examined studies ranked, until recently, from 13 analysed newspaper articles (McLeod and Hertog, 1992) to 705 (Di Cicco, 2010). This is likely due to the time and resource-intensive task of manual coding for content analysis or qualitative reading. Only recently, 5 studies have introduced larger samples and more sophisticated sampling methods (Corrigall-Brown and Wilkes, 2012; Harlow et al., 2020; Hughes and Mellado, 2015; Lee, 2014; Wouters, 2015a). Hughes and Mellado (2015), who use the largest sample, analysed 7,386 newspaper articles over a 21 year period employing the constructed week method.[3] In the last decade, methods to significantly reduce the resources needed to code texts by employing machine learning and other computational techniques for content analysis have become more advanced and are more regularly employed in the social sciences (see Grimmer and Stewart, 2013, for an overview). They allow to pass on sampling completely, therefore, increasing efficiency of the analysis and decreasing uncertainty (e.g., King et al., 1995, pp. 66-74). However, to the best of my knowledge, these methods have not been employed to assess protest coverage. Doing so thus makes findings generalisable to a greater degree, which will be one of the main contributions of this thesis.

Furthermore, many of the studied cases are relatively dated, especially the seminal work that informed the establishment of the *protest paradigm* as a concept. Gitlin (1980), Halloran et al. (1970) and Murdock (1973) all published their studies about

[3] One Monday, Tuesday, Wednesday, et cetera is randomly selected from each year.

movements in the 1960s. Since then, media landscapes, protests and society have changed considerably. As will be discussed in Section 3.3, this is thought to have a considerable impact on news media coverage of protests.

We also still know relatively little about how diverse cultural influences on journalists in different countries shape adherence to the *protest paradigm*. Most early studies that supported the idea of a *protest paradigm* focused on the US, while results outside the US are more sceptical. Wouters (2015b), for example, found that Belgian media usually highlight the issues protests seek to address. The patterns derived from the *protest paradigm* also appear less salient, to varying degrees, in protest coverage by Brazilian, Chinese, and Indian news media (Shahin et al., 2016). In a direct comparison, Dardis (2006b) established in a cross-national comparison between Iraq War protests' coverage in the US and the UK that socio-political differences between countries condition the use of the paradigm: the US media followed the paradigm more closely than journalists in the UK. Filling this gap with more comparatives studies would be necessary to generalise findings beyond country borders.

Finally, what drives the adoption of the *protest paradigm* is not entirely clear. There are several theoretical approaches and a number of factors have been suggested to lead to reporting as described above. These are discussed and further developed in Section 3.4.

Chapter Conclusion

This chapter has surveyed the theoretical and empirical contributions and gaps of previous research that looked at the selection and description of protest by the mainstream media. Ultimately, literature on the study of protest in the news can be characterised by two generalisations: there is consensus about the existence of a selection bias of protest events in the news and there is considerable evidence that a description bias — mostly referred to as *protest paradigm* — is prevalent in reporting about protest.

However, while there has been plenty of research on the *protest paradigm* over the years, existing studies also show limitations in some respects and their findings appear contradictory in others.

Specifically, most research so far employed relatively narrow case studies that focused on a specific, often radical, protest event or a number of events over a short period of time. There has been a lack of longitudinal studies which analyses how the paradigm evolves over time and varies across issues. Furthermore, studies are sometimes hard to compare as they employ different operationalisation of the paradigm and ways to measure it. Since many earlier studies do not quantify the coverage that follows the paradigm, it is impossible to assess their allegation of dominance or compare it to other points in time. Especially since the theoretical foundation of the paradigm does not provide guidance on the interpretation of results: there is no clear cut-off point at which coverage is seen as following the paradigm. The assessment thus seems somewhat subjective to the researcher, with two studies arriving at different conclusions while assessing the same material (Edgerly et al., 2011; McLeod, 2007). The following chapters will address these gaps in our knowledge about protest coverage, which I aim to fill through the theoretical, methodological and empirical contributions this thesis.

Chapter 3

What Makes the Media Frame Protest?

The last chapter has introduced the arguably most central concept in the study of media coverage of protest — the *protest paradigm* — and surveyed previous studies which employ it. What became apparent is that while the paradigm has been used in several dozen studies, there is some disagreement on how it is conceptualised. Furthermore, while it was originally assumed that the marginalisation devices described in the *protest paradigm* made up the overwhelming part of reporting about protest, recent studies have cast doubt on this idea.

This chapter aims to provide a solid theoretical underpinning for the study of protest coverage. For that reason, Section 3.1 discusses the framing concept, which allows to systematically analyse large quantities of media data and link the outcome to theories of framing effects. Framing is, therefore, used as a theoretical lens through which media content is analysed and made sense of. Section 3.2 discusses the influences which shape newsmaking and news content. Section 3.3 then asks what can be expected of media coverage on protest over time. Some of the studies which found the *protest paradigm* less prevalent than previously assumed speculate that this might be due to changes in the media ecology and broader trends in society. These changes are hence discussed. A second explanation why findings on the existence or prevalence of the *protest paradigm* are more mixed than in studies in the past is that the logic with which it is applied might be more conditional than originally assumed. Moreover, as much of the seminal work in the field studied specific cases, these possibly had some rather specific features that amplified the use of the *protest paradigm*. Section 3.4, therefore, covers the factors which might lead to more or less prevalence of de-/legitimising coverage.

3.1 Media Framing

One problem identified in the last chapter is that there is theoretical gap in the *protest paradigm*: it does not provide a clear strategy on how to interpret results as the normative assumption that coverage following the paradigm will hinder a protest's chances to success does not spell out under which circumstances this is supposed to happen. I turn to the framing concept as a suitable theoretical lens through which to analyse media content. Framing is arguably the most often employed concept in communication and media research (as highlighted by overviews such as Borah, 2011; D'Angelo, 2002; D'Angelo and Kuypers, 2010; Entman et al., 2009; Scheufele, 1999). Importantly, McCurdy (2012) found that the bulk of studies concerned with media coverage of protest employ some form of framing research. The main theme of the framing concept is, in short, that in order to make sense of the seemingly meaningless succession of everyday life's events, people select and organise certain aspects of what is happening into consistent frames (e.g., de Vreese et al., 2001; Goffman, 1974). More specifically, in communication, to frame means to pre-select and emphasise some information while disregarding other in order to tell a coherent story (Wolfsfeld, 2011). In doing so, communication actors, such as mass media, can affect how the audience perceives, understands and remembers issues and, consequently, how people evaluate and choose to act upon them (Entman, 1993).

The two big strengths of framing for this thesis are: first, the concept provides a theoretical lens through which to analyse media content and condense a plethora of stories into meaningful categories; and second, framing promises to bridge several research areas such as the production, content, and effects of news (Matthes, 2009). However, beyond being a concept, there is some significant discordance among scholars whether framing is an approach (e.g., Pan and Kosicki, 1993), an analytical technique (Endres, 2004), a theoretical tool (Matthes, 2009), a theory (e.g., Scheufele, 1999), a (fractured) paradigm (e.g., Entman, 1993), or a multi-paradigmatic research program (D'Angelo, 2002).

In the study of protest, framing research broadly falls into two categories, following a general divide in framing research: following McLeod and Hertog (1999), most studies employ a strategy that aims at distilling narratives and story devices from protest coverage to describe different types of reporting as frames. Other authors employ a concept introduced by Iyengar and Kinder (1987) to distinguish between the meta categories *thematic* and *episodic* coverage. *Thematic* framing contextualises an issue or problem. It provides information about the general development of a raised issue or conditions that may cause it, often providing perspective on how widespread a problem is. *Episodic* framing, in contrast, discusses only a concrete case, which often renders an issue as an individual's problem (Iyengar, 1991). Protesters usually seek thematic coverage, as their main goal is to raise awareness of a problem or certain aspects of it. Episodic coverage, on the other hand, might focus just on the fate of an individual protester who is affected by the issue or even just describe the protest event itself, reporting just crowd sizes and public disturbance. Some authors, therefore, conceptualise episodic coverage as adherence to the *protest paradigm* (Jha, 2007; Smith et al., 2001). The understanding of frames as *thematic* and *episodic* coverage, however, is overall less well linked to the majority of findings about the *protest paradigm*.

In this thesis, I employ the most common and one of the richest definitions of framing:

> To frame is "*to select some aspects of a perceived reality and make them more salient in a communicating context, in such a way as to promote a par-*

> *ticular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation* for the item described" (Entman, 1993, p. 52, original emphasis; for a comparison of framing definitions see Matthes, 2009).

Using this definition makes it possible to study a large number of media reports by extracting a relatively small set of features from each text and distilling these further into frames (see Section 4.2.3).

However, why is it important to assess framing patterns in the coverage of protest? As said above, the framing concept bridges several research areas, importantly the content and effects of news. In Section 2.1, I mentioned that the media selection bias is linked to *agenda setting* effects for protest: if the media do not pick up a protest, it is unlikely that it will be discussed in the political discourse. Likewise, the reason to study the media description bias embedded in the *protest paradigm* is the expectation of framing effects, namely the delegitimisation, marginalisation and demonisation of protests (McLeod and Hertog, 1999).

A bulk of attention in framing research has been on how different frames in communication influence audience frames and individuals' opinions (Chong and Druckman, 2007b). This research can be useful when making sense of framing patterns, which is the goal of this thesis. Specifically, framing effects research asks if and under which conditions individuals focus on the considerations emphasised by framing while constructing their own opinions (Druckman, 2001). Assuming that these effects occur unbounded, it would make the media an incredibly powerful actor. They could manipulate what people believe to be the most important considerations about an issue and thus greatly shape public opinion (Entman, 2007). In fact, this is what earlier empirical evidence suggested (e.g., Cappella and Jamieson, 1997; Price et al., 1997) and was notably concluded by McLeod and Detenber (1999) to be the effect of *protest paradigm*-frames. However, if opinions could be determined so easily, we would see much more substantial and generalised media effects. Moreover, it would cast serious doubt on the capacity of citizens to participate in the democratic process, as they would merely be able to regurgitate opinions provided to them by the media.

Albeit, these studies usually draw their conclusions from simple designs that expose two or more groups to news reports with one dominant frame per group leading to the predictable result of individuals employing the one frame they were exposed to in their considerations (Chong and Druckman, 2007a; Matthes, 2013). Studies that employ more advanced designs found that under realistic circumstances — a political debate with more than one perspective — framing effects tend to cancel each other out (Sniderman and Theriault, 2004) or at least complicate the cause-effect relationship substantially (Chong and Druckman, 2007a; Entman, 2010).

If many of the studies about the *protest paradigm* operated under the assumption of unbounded framing effects, it would explain why they often did not assess the prevalence of *protest paradigm* reporting: even a single article adhering to the paradigm might have dramatic consequences. Furthermore, an implicit difference between authors expectations about the effects might explain the different assessments in *protest paradigm* studies highlighted in Section 2.2.2.

Although not the focus of this thesis, for theoretical clarity, it makes sense to ask: under which circumstances are framing effects expected to appear and why? Again, there is no lack of theory in the framing literature to answer this question, but rather a confusing surplus of theoretical and meta-theoretical approaches (e.g., D'Angelo and Kuypers, 2010; De Vreese, 2005; Entman, 1993; Pan and Kosicki, 1993; Scheufele, 1999). For this research, the theoretical effort made by Chong and Druckman (2007c) seems most fruitful: they integrate existing approaches by tracing the psychological mechanisms behind framing effects and furthermore include factors like competition and strength of frames into their model. Building on the *conventional expectancy value model*, Chong and Druckman (2007c) consider an attitude of an individual towards an object as a combination of the evaluative beliefs of that person on a dimension and the salience of that dimension for evaluation. For example: a person may regard an issue as advantageous on an economic dimension and harmful on a social justice dimension. If the economic dimension appears more salient to that person, she can be expected to have a positive opinion towards the issue.

Framing effects can thus result from introducing new beliefs to an individual's overall attitude as long as the recipient accepts and prioritises the new considerations while constructing her opinion. Alternatively, existing attitudes can also be changed by either altering an individual's beliefs — which is referred to as persuasion — or by altering the salience of one dimension. In order for this to happen, framing effects would have to increase the *availability* (a consideration must be stored in an individual's memory and the individual must comprehend its meaning), *accessibility* (the consideration must come to mind when reflecting on an issue), or *applicability* (the consideration must be judged relevant to the issue) of certain considerations (Chong and Druckman, 2007b,c). From this explanation, several expectations can be derived: The most obvious one is that frequent repetition of a frame can ensure its availability and increase its accessibility — an explanation for framing effects that was given by several political communication scholars (e.g., Cappella and Jamieson, 1997; Iyengar, 1991; Price et al., 1997). However, it leaves aside the applicability of considerations. For example: demonstrations usually interfere with traffic. Framing protests as a nuisance for motorist again and again will probably not lead people who do not drive to judge a protest negatively. And in fact, newer experiments showed only minor effects of frequent repetition and no effects whatsoever if the repeated frame has strong competition (Chong and Druckman, 2007a).

In non-competitive environments, individuals may use whatever considerations are made available and accessible to them. But under other circumstances, only strong, compelling frames can influence people. What exactly constitutes a strong frame — apart from being persuasive — is not clear, though. Yet, what has been established is that as soon as multiple interpretations are available, individuals tend to deliberate and personal values and evaluations of the quality of arguments become more important (Chong and Druckman, 2007a; Schemer et al., 2012). Even after individuals accepted a frame, it is likely to be discarded if it is weak and they hold their opinion with low certainty (Matthes and Schemer, 2012). Furthermore, some people are more knowledgeable and motivated than others and are, therefore, more prone to evaluate the applicability of frames. Yet, even unmotivated and less knowledgeable individuals have been found to choose strong frames in competitive environments (Chong and Druckman, 2007a).

For the state of knowledge of protest coverage, this suggests a significant limitation: The emphasis on the sheer presence of the *protest paradigm* might have led to overemphasising the importance of the phenomenon. As long as relevant information is still salient enough to make it available and accessible to the audience, it is plausible to assume a protest, even in the presence of some dimensions of the paradigm, can be successful in the sense of drawing attention to an issue and sparking a public conversation about it. Only when the elements of the *protest paradigm* form a frame that completely dominates the discussion, strong delegitimising framing effects can be expected.

For this thesis, framing thus provides a theoretical framework to assess potential results: will I find a situation in which multiple framings of protest events are available and accessible for the public? In this case, the conclusion will be that the audience will choose for themselves which of the competing frames they find most convincing. Or will I find that one or several frames dominate the discussion completely? In this case, one might expect strong framing effects — such as the delegitimisation, marginalisation and demonisation of protests. This calls for a quantification of results to assess how prevalent or even dominant certain frames are. Furthermore, a different consideration regarding the framing concept also suggests that a simple distinction between coverage that follows the *protest paradigm* and coverage that does not is ill-advised: Dardis (2006a) finds that even distinctly protest-supportive coverage makes use of at least some *marginalisation devices*. This is not surprising given that an article can contain multiple co-existing, overlapping or competing frames. A study that works with only two broad categories might easily miss such subtle distinctions.

Identifying and quantifying the main frames in the coverage of protest in the UK is thus a prime goal of this thesis. Consequently, the first research question is:

> **RQ1**: How do British newspapers frame the coverage of domestic protest events?

3.2 Random Reactions to Random Events?

> "Our product is put together by large and shifting groups of people, often in a hurry, out of an assemblage of circumstances that is never the same twice. Newspapers and news programmes could almost be called random reactions to random events. Again and again, the main reason why they turn out as they do is accident — accident of a kind which recurs so haphazardly as to defeat statistical examination" (quote from an interviewed reporter in Murdock, 1973, p. 163).

The major premises of researching frames in communication are that there are always different perspectives a frame could potentially reflect and that, contrary to the belief of many journalists, media content is not the product of random reactions to random events but shaped by broader forces. The notion that media content is influenced by the environment and circumstances of newsmaking often upsets journalists who claim to follow ideals of professionalism, such as objectivity, and mostly try hard to only report the facts as they occur in the "real world" (Schudson, 1997). However, the mirror hypothesis — the idea that media could reflect reality if journalists just successfully avoid bias — has been broadly rejected (e.g., Bennett, 1996; de Beer, 2010; Hamilton, 2004; Shoemaker and Reese, 2014). The reason is that it is seen as problematic to "think of a reality out there with which we can compare mediated content" (Shoemaker and Reese, 2014, p. 3).

A simple thought experiment might help to explain this idea in the context of protest. Imagine a protest participant with a clear idea of what the protest is about and who experiences peaceful protest in their immediate surrounding. Now think of a police officer sent to the rowdiest section of that same protest march to establish public order. Now think of a bystander or commuter or owner of a shop along the march route inconvenienced by the protest. And finally, think of a reporter who was sent to the event to report about it. Maybe the reporter gets a good look at the crowd size from an elevated position. They might ask several different participants why they think the protest is necessary. They might talk to the police, officials and experts on the protest issue. Yet, is the reporter's perspectives any more representative of reality than the account of the other described people?

Media are often criticised for allegedly providing a biased or misleading account of reality. However, as Schudson (1997) reminds us: "journalists write the words that turn up in the papers or on screen as stories [...] not 'reality' magically transforming itself into alphabetic signs" (1997, p. 141). As journalists are human beings, they cannot lift themselves of their "human context and apprehend reality apart from it" (1997, p. 3). Or, put differently, news represent the world but cannot mirror it. Since human beings must select information in order to tell a coherent story, some people must make the decision what to present and how to present it (Schudson, 1997, p. 33).

Additionally, not only must journalists try to craft a compelling and informative story, but they also work against the clock and have the time or page-space constraints of their outlet in mind. Individual journalists, as well as media organisations, are therefore eagerly "routinizing the unexpected" (Tuchman, 1973, p. 110) by creating practical and effective procedural rules and routines (Gans, 1980; Hirsch et al., 1977; Shoemaker and Reese, 2014). So despite changing circumstances and varying details in each new story, many scholars have argued that formal constraints as well as the norms and standards of their profession drive journalists work to such an extent that homogeneity, within and among outlets, should be expected to some extent. This indeed has often been observed in empirical studies (e.g., Cater, 1959; Cook, 2005, 2006; Hirsch, 1977; Sparrow, 1999; Zaller, 1999).

Since the 1970s, studies in the sociology of news production and in other fields have tried to uncover the principal forces that shape the news. From this body of work, several aspects are relevant for the study of protest coverage. These aspects are introduced in the following sections as they are linked to how media coverage is expected to change over time (Section 3.3) or why certain events receive different coverage than others (Section 3.4). Most of the time, it is expected that newsmaking is structured around *news values*, which represent what journalists and gatekeepers, such as editors, regard as "newsworthy" (Shoemaker and Reese, 2014, pp. 170–173). These values are in turn shaped by assumptions about what audiences will find appealing. Different scholars have attempted to map these values and have compiled different lists. Nevertheless, it is apparent that news values underpin the assumption that the *protest paradigm* is prevalent.

Shoemaker and Reese (2014), as a prominent example, provide that *prominence and importance*, *conflict and controversy*, *the unusual*, *human interest*, *timeliness* and *proximity* are the features people find interesting to know about. *Importance* in the case of protests might highlight if there was property damage or injuries during an event rather than other issues the protesters aimed to highlight; a focus on *conflict and controversy* likely gives precedence to clashes rather than messages; highlighting *the unusual* underlines the assumption that protesters who appear odd and different from mainstream society will be heavily featured in reporting; and *timeliness*, or the assumption that "people have limited attention spans and want to know what is happening now" (Shoemaker and Reese, 2014, p. 171), would explain why reports focus on the method over a potentially complicated message of a protest. Other compilations of new values also highlight aspects such as surprise, drama, negativity, or deviance, which are considered more newsworthy (Boydstun, 2013; Galtung and Ruge, 1965; Hetherington, 1985). Ultimately, the goal of news outlets is to make their product attractive (Hamilton, 2004), which they try to do by catering to perceived audience demands (Gans, 1980), using cultural cues (Bird and Dardenne, 1988).

From a broader theoretical view, the existence and prevalence of the *protest paradigm* is, therefore, plausible. And as mentioned in Section 2.2.2, the majority of studies has concluded just that. Therefore, the first hypothesis for this thesis is:

- **H1**: Stories on protest events mainly use delegitimising framing as described in the protest paradigm literature.

However, considerations of newsworthiness change over time and protest events have widely differing features. So what specifically can be expected when studying protest over a long period of time and when considering a collection of arguably very different events? This is the focus of the next section.

3.3 Changing Media Ecology, Changing Protest Coverage?

One of the issues identified in Section 2.2.2 as a shortcoming of previous research on protest coverage was the limited time frame studies employed. But why is this question important? In short: because a lot has changed in the last decades. Specifically, scholars have argued that there has been a transition from a low to a high-choice media environment in the UK and other countries, which led to more competition among legacy outlets and with emerging information channels on social media sites and the rest of the internet (van Aelst et al., 2017). These changes coincide with developments in consumer demands monitoring, more opportunities for protesters to interact with the media as well as larger audiences on their own, and the steady normalisation of protest in society.

However, if the media and society changed in the time frame of this research (1992-2017), did protest reporting change with it? This leads to the second research question:

RQ2: How — if at all — did the framing of protest reporting change over the last 26 years?

In the time frame of this research (1992-2017), the principal catalyst for change in the media ecology were new technologies: First, satellite television altered the structure of the broadcasting market in the 1990s and substantially changed consumption patterns as the number of TV channels the audience could choose from grew rapidly from three to several hundred channels (Negrine, 2016). Noteworthy is also the introduction of rolling news channels which had the significant competitive advantage when it came to breaking news, with rapid responses to ongoing events and 24/7 of broadcasting to fill with content (Kuhn, 2007; Stanyer, 2010). Since the turn of the millennium, however, media environments have undergone yet another and even more drastic transitional phase (e.g., van Aelst et al., 2017): an entirely new media sector — online news — has emerged since the mid-1990s and started to gain traction with the beginning of the 21st century (Just, 2013). Only a few years later, online news outlets were additionally joined by private internet blogs and social media, which made it possible for citizens to comment on current affairs or provide news on their own. The internet, furthermore, diminished national boundaries for news access, which means that British news outlets compete with outlets from around the globe now — at least in theory (Curran et al., 2013).

The explosion in the number of news sources has led to an unprecedented amount of information available to audiences by a multitude of outlets that was formerly unthinkable, leading to a public sphere "larger, denser, and accessible to more people than at any previous point in Britain's cultural history" (McNair, 2006, p. 39). On top of that, news providers also compete for audiences' attention with an ever-growing amount of non-political information and a host of leisure pursuits, leading to what van Aelst et al. (2017) call "high-choice media environments" (p. 4).

For the British case, it should be noted that the most important technological change, the penetration of internet use, started early in the UK and lagged behind the US less than the bulk of continental European countries in the 1990s (see Figure 3.1). British media organisations were, furthermore, early to actively use the internet to gather information and provide online news services. By the 2001 general election, online information and communication about politics were dominated by the established news organisations, though the tabloid press lagged behind the efforts of broadsheets and the BBC (Coleman, 2001). Kuhn (2007) remarks that fear of losing advertisers and audiences to online media as well as presenting themselves as “the vanguard of forces embracing technological change” (p. 18) brought UK newspaper outlets to move online early.

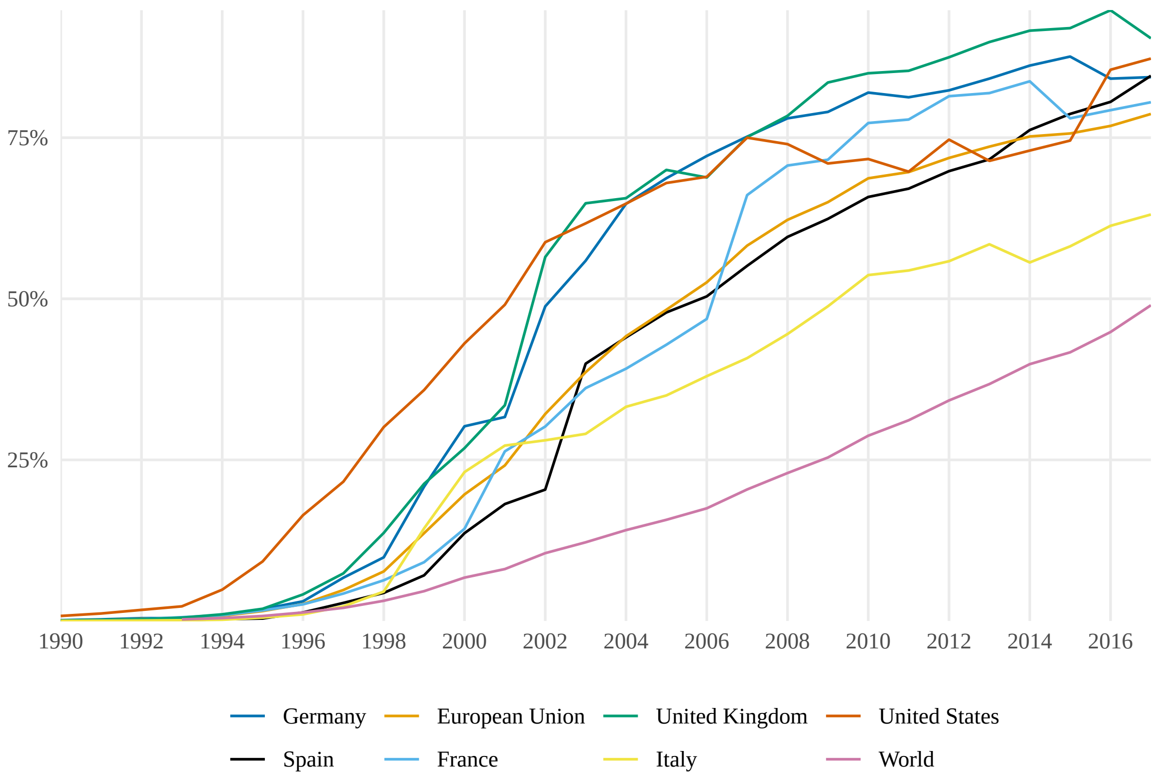


Figure 3.1: Individuals using the Internet (percent of population), Selected Countries (Data Source: World Bank)

These technology-driven changes have several potential implications for this thesis, which focuses on newspaper content: Firstly, in the increased competition the added supply of news created, the newspaper industry is in a constant retreat since its high-point in the 1950s — in the Liberal countries, including the UK, even more so than in continental Europe (Kuhn, 2007, pp. 7-9; Hallin and Mancini, 2004)[4]. Thus, the newspaper industry operates in a market where economic pressures keep growing as the sector competes for a shrinking market. Consequently, the shrinking revenues have already forced a considerable number of British local newspapers out of the market (Cox, 2016). Economic incentives are therefore expected to grow in the time frame of this thesis.

However, the effects of growing economic incentives on reporting are not yet clear (see van Aelst et al., 2017). On the one hand, economic incentives are thought to pull journalists into the direction of less critical reporting as dependencies towards elites and advertisers grow (Herman, 1995), and diminished resources in journalism take their toll. This means that the influence of corporate and political elites' strategies to influence the media via information subsidies and other PR techniques likely grows over time (e.g., Davis, 2000, 2003; Gandy, 1982). However, it must also be noted that a loss in journalistic reputation can quickly lead to a shrinking audience and hence financial losses (Hallin, 1986a; Schudson, 1997). As competition increases, providing patently misleading information or failing to cover an angle the audience wants to know about can thus directly and quickly reduce an outlet's revenue as consumers turn to the competition. Political goals of media owners, who concentrate an increasing amount of power, might nevertheless trump these concerns, as has been shown for Rupert Murdoch's influential *News Corp*, which owns leading and often explicitly partisan outlets in several countries (e.g., McKnight, 2010; see also van Aelst et al., 2017). Furthermore, this effect depends strongly on what the audience actually wants — accurate reporting or confirmation of their own views. Different co-existing audience demands might thus be the reason why empirical evidence for decreasing levels of

[4] More recent figures show that this trend continues: pressgazette.co.uk/national-newspaper-print-abcs-for-jan-2017-observer-up-year-on-year-the-sun-is-fastest-riser-month-on-month/.

critical and diverse reporting is mixed (Aalberg et al., 2016; Reinemann et al., 2016; van Aelst et al., 2017).

Secondly, said audience demands have become more important and also more palpable since the 1990s: As mentioned above, financial imperatives dictate that outlets should try to provide information audiences are interested in (Hamilton, 2004). This has never been more important than in today's media ecology in which content is available everywhere and all the time (Ksiazek et al., 2014). Yet, it has also never been easier to gather user metrics that provide more insights into user demands than ever before. A growing body of literature thus shows that editors become more aware of and amenable to consumer tastes.[5] In the pre-digital world, there had just not been any satisfactory methods to retrieve audience feedback (Schlesinger, 1987), although market research and letters to the editor allowed sporadic insight into what the public deemed important (Wahl-Jorgensen, 2002). Yet, "in the twenty-first century the audience has invaded the newsroom" (Just, 2013, p. 107).

When comment sections and user metrics became widely available, editors started to get a clearer sense of what the audience demands and have been shown to take this into account (e.g., Anderson, 2011a,b; Boczkowski and Peer, 2011; Dick, 2011; Loosen and Schmidt, 2012; MacGregor, 2007). What this means for news content, however, is not entirely clear. Often it is assumed that audiences generally tend to favour "soft" news while journalists' preference of "hard" news (public affairs) has led to a focus on these topics in the past (e.g., Boczkowski, 2010; Boczkowski and Peer, 2011). This could mean that reports about the messages of protest become less frequent and, crucially, less detailed over time if they are seen by audiences as "hard" news and are thus scorned. However, while the decline in "hard" news is a popular claim, actual evidence is rather mixed (Reinemann et al., 2012; van Aelst et al., 2017). Furthermore, if detailed stories about some protests are in high demand with the audience — as

[5] The trend is nicely illustrated when comparing the different findings by Quinn et al. (2002), MacGregor (2007), Peters (2010) and finally Vu (2013) over time: each study attests a stronger influence of audience metrics than the last.

could happen if a grievance voiced by protesters is widely spread or widely polarising among audience members — it can be expected that outlets allocate more resources towards its coverage. In short, consumer demands are assumed to shape news content to a much larger degree than in the past — yet it is not clear if for the better or the worse.

Third, more outlets supposedly means more perspectives on the same event or story. The historical monopoly of newspapers as the gatekeepers to what stories make it on the public's agenda has finally been thoroughly broken. Today's media landscape can better be described as "hybrid media systems" (Chadwick, 2017), in which an increased number of actors, namely in broadcasting, online news and social media sites, can "introduce, amplify, and maintain topics, frames, and speakers that come to dominate political discourse" (Jungherr et al., 2019, p. 409). Issues may start and/or gain traction on social media, which then catapults other media types into "storm mode" (Boydstun et al., 2014), that is, explosive and sustained attention (Langer and Gruber, 2021; Singer, 2014). Additionally, people perceive issues as more salient when exposed to them on social media (Feezell, 2018), increasing audience demand for further updates.

The hybrid media system could play out differently in different media sectors. For online news, it was shown that the event-driven character provides far more diversity in reporting compared to established media outlets (Humprecht and Büchel, 2013). Newspapers, however, have also moved online to a large extend and provide nearly immediate coverage as well. According to Bennett (2003), event-driven reporting can mean that journalists break with some norms and routines to insert their own views until elites and officials can get a handle on the press. For protest news, this was illustrated by Tenenboim-Weinblatt (2014). On the other hand, as the objectives of a protest takes time to research, a quick media response could focus stronger on event characteristics — especially if a protest group is new and therefore unknown (Wouters, 2015b). Furthermore, Curran et al. (2013) conclude that "the incumbent economic power of leading media conglomerates, and the constraints exerted by the wider social context across all media" (p. 893) lead the most popular online news websites to report in a very similar way to their established counterparts offline: voices of authority and expertise are still favoured over ordinary citizens and activist organisations; news still have a spatial bias with a strong focus on the home nation; the selection of stories closely resembles those in broadcasting and the press. For social media, empirical results are again somewhat mixed. Diversity of different views is undoubtedly much greater on social media, yet most social media accounts receive little attention. Overall, the *protest paradigm* plays a reduced role on social media but can still be found regularly (Harlow and Johnson, 2011; Harlow et al., 2017). Furthermore, evidence from the US suggests that the paradigm is also still common among journalists who cover protest via social media (Araiza et al., 2016; Harlow, 2020), although the same was not found in a Brazilian case study (Mourão and Chen, 2019).

Fourth, as the working conditions for journalists changed over time, so did the circumstances under which protesters operate: The internet was argued to be a tool for previously excluded groups to participate in the public debate, which is a part of the political process (Poster, 2001). For protesters, setting up an outlet on the world wide web has had marginal costs that do not remotely compare with the necessary expenses to set up a newspaper or broadcasting channel for a while already. Since social network sites have become a mass phenomenon, however, those costs have been reduced to basically zero (boyd, 2011; Enjolras et al., 2013). For protest organisers, this means that they can use these relatively open public spheres to discuss ideas, build networks, mobilise participants, plan events and — crucially — communicate with journalists and a wider public.

Bennett (2003) has argued that these new means, accompanied by other developments generally subsumed under the term globalisation, have led to a new form of globally operating protest networks that first took shape by the late 1990s (also see Bennett, 2010). Theoretically, the new environment in which both protesters and journalists operate enables both to work together more quickly. Journalists now heavily employ social media — especially Twitter — to gather information, communicate with colleagues and distribute news in order to scoop competitors (Parmelee, 2013). That means ideas spread by low-resourced organisations or even individual protest organisers can serve as new material for journalists who try to deliver timely updates on unfolding events (Enjolras et al., 2013; Ismail et al., 2019; Shirky, 2009; Theocharis et al., 2014). Evidence suggests that journalists derive new ideas for stories from social media sites (O'Connor, 2009; Parmelee, 2013; Solis, 2009) and content from social media users is used as sources in the mainstream press (Bennett and Segerberg, 2012; Ismail et al., 2019; Stanyer, 2010). Especially when no other content is available (yet) (Bennett and Segerberg, 2015; Just, 2013; Oz, 2016). In turn, journalists themselves are also much more accessible today since personal email addresses and public accounts on social media make it easy for everyone to contact them directly. Furthermore, protesters are becoming more aware of how the mainstream media operate and the opportunities to influence media coverage through PR techniques (Cammaerts, 2012). However, a direct communication line between protesters and journalists represents an ideal scenario that describes only a fraction of media-savvy or lucky protest groups (e.g., Wolfsfeld et al., 2013). Nevertheless, these developments could represent a potential shift in power from elites to digitally mediated collective action (Bennett and Segerberg, 2012).

Finally, society itself has also changed considerably over time. This is important as the narratives, stereotypes and myths reporters employ to craft their stories are specific to the broader culture at a particular time and age: "The media operate within the culture and are obliged to use cultural symbols" (Schudson, 1997, p. 151). Di Cicco (2010), for example, suggests that after decades of adverse reporting on protests, the image of protests as a nuisance might gain foothold in society and thus influence future generations of audiences and journalists. In contrast, most other authors suggest that since the 1960s, when seminal work first explored the portrayal of protest in the news, there has been a considerable process of normalisation of protest as a legitimate political tactic (Dalton, 2004; Sanders et al., 2003; Norris, 2002, pp. 188-212; van Aelst and Walgrave, 2001; Oliver and Maney, 2000). Today, many different groups employ protest as a tool, and many more people today have participated in a protest at least once in their lives (e.g., Borbáth and Gessler, 2020; Hutter and Borbáth, 2019). Furthermore, there is little difference today between demonstrators and the demographic in the electorate at large, suggesting that demonstrating has become mainstream (Cottle, 2008; Norris et al., 2005).

Consequently, protest is no longer just a phenomenon of left-wing groups, as had primarily been the case in the 1960s and 1970s. Weaver and Scacco (2012), for example, observed that the radical right-wing *Tea Party* in the US copied many of the protest tactics the left had used decades earlier. In turn, right-wing media were much more sympathetic towards those protests. In Britain, Milne (2005) provides some observational evidence that protest cannot even be pinned down on a right-left spectrum any longer as events started to address single issues such as rising petrol prices or protests for and against a ban on fox-hunting. The press is assumed to find some protest objectives more acceptable than others. This might explain why Milne (2005) also finds that sections of the British press frequently get behind or even champion specific protests that are in agreement with their editorial lines. Thus, the normalisation of protest is likely to continue. Probably along with the normalisation of protest coverage.

Summing up, since the protest events of the 1970s that were documented by the most seminal studies in the field (i.e. Gitlin, 1980; Halloran et al., 1970; Murdock, 1973), there were substantial developments in the media landscape, the circumstances under which journalists and protesters operate and the larger societal context of protest reporting. As the industry changed, there has been considerable discussion about if and to what degree existing findings about journalism would persist over time (Cottle, 2003). Although the evidence about the influence of these developments is not entirely clear yet, the suggestion by authors like Cottle (2008) that old findings about the characteristics of protest portrayal in mainstream news media must be revisited seems more than plausible. Some of the described developments, such as increased economic pressures and closer monitoring of audience demands, seems to point in a direction in which "soft" news and superficial coverage gain in importance over time. Yet overall, it is assumed that "today's media ecology arguably contains more political opportunities for dissenting voices and views from around the world than in the past" (p. 859) (also see, e.g., Rohlinger, 2020; van Aelst et al., 2017). Consequently, the evidence we have about the influence of developments over time suggest that the *protest paradigm*, which was "theorized within a media ecosystem that was more industrial than digital news audiences encounter today" (Kilgo et al., 2019, p. 418), might be applied more conditionally (Harlow et al., 2017; Kilgo and Harlow, 2019; Wouters, 2015b), and less frequently over time (Araiza et al., 2016; Kyriakidou and Olivas Osuna, 2017; Lee, 2014).

For this long-term empirical study, the theoretical expectations are thus that:

- **H2a**: Delegitimising framing decreases in salience over time.
- **H2b**: Legitimising framing increases in salience over time.

3.4 What Drives Protest Coverage?

Most of the evidence we have regarding the factors which shape the conscious or unconscious choice of frames by the news media when covering protest comes from the strands of literature mentioned in Section 2.2. As stated, there is a long tradition in the sociology of news production and other fields that try to uncover principal factors that shape the news. However, what became clear is that the task is not an easy one as "the list of possible variables is almost endless" (de Beer, 2010, p. 189). Shoemaker and Reese (2014), as arguably one of the most prominent examples of the field, propose to divide factors into five distinct levels, ordered from micro to macro: The *individual* level, which describes the characteristics of the individual journalist or editor; the *routines* level, which describes the immediate constraining and enabling structures and routines within journalists work; the *organisation* level, which comprises the larger context of the routinised activities such as the goals, policies and structure of an organisation; the *social institution* level, which comprises influences on media types, such as the press, or other subsets of the media landscape; and the *social system* level, which describes influences on content from the system as a whole. What this thesis is most interested in are factors on the level of routines, organisations and the broader social system, which are thought to undergo several changes in the selected time frame.

Based on the existing literature, both the broader sociology of news and previous studies about protest coverage, I formulate theoretical expectations in this section regarding the third research question:

> **RQ3**: Which factors explain the choice of frames by the news media when covering domestic protest events?

While the concept of testing factors which condition protest coverage is not new (e.g., Boyle et al., 2012; Kilgo and Harlow, 2019; Lee, 2014), the set of variables presented in this section is novel and, to my knowledge, the most complete so far. To make discussion of the factors easier, I follow Shoemaker and Reese (2014) by dividing them into levels, ordered from micro to macro. However, the three levels I use are different as they are closer tailored to the specific topic at hand — finding the influences on how *protest* is reported. Specifically, the factors taken into account here are divided into event-level, outlet-level and factors connected to time.

Note that at this stage, it is unknown which frames the analysis will reveal, making it impossible to specify which frames might correlate with which factors. Many previous studies worked with the dichotomy of coverage adhering to the *protest paradigm* and coverage that is more supportive of a protest goal — which I will call *delegitimising* and *legitimising* coverage in the following. As it is expected that most frames can be placed within this broad dichotomy, the hypotheses in this section capture the expected relation between factors and a potential set of frames that are, most likely, legitimising or delegitimising. In other words, while the analysis allows the revelation of more fine-grained frames, this section will discuss the theoretical expectations for relationships between the factors with legitimising and/or delegitimising coverage of protest. However, the dichotomy of legitimising and delegitimising coverage is unlikely to tell the entire story. What aspects of an event do news highlight and which do they omit specifically? Depending on the specific framing, the degree to which reports legitimise or delegitimise a protest cause could vary substantially. So while the described dichotomy is useful to discuss factors that drive reporting, the ultimate goal is to measure the influence these factors might have on the usage of specific frames.

3.4.1 Event-Level Factors

Protest Goals. Several studies document that the goals or causes voiced by protesters shape the coverage these events receive. These studies are based on the idea that the media are agents of social control who consciously and unconsciously reward conformity to and punish deviance from a perceived mainstream or status quo (McLeod and Hertog, 1999). In this regard, Shoemaker (1984) found that groups that are perceived as more deviant or radical tend to receive more delegitimising coverage (Hall et al., 1978)[6]. Based on this, Boyle et al. (2004) tested how the *level of deviance* of protests shape their coverage. They conclude that coverage of deviant protest conformed closer to the *protest paradigm*, conceptualised as more critical coverage, coverage employing episodic framing and coverage that is less likely to use protesters as sources.

How this might work in practice was provided by Tenenboim-Weinblatt (2014), who studied in in-depth interviews how journalists perceived their hand in framing a social protest in Israel. At first, individual journalists were supportive of the movement as they identified strongly with protesters, who belonged to a similar socio-economic class, and their message, which criticised the same issues reporters faced in their daily lives. Yet, journalists noted that this identification and sympathy for the protest demands

[6] Also see the deviance sphere in Hallin (1986b).

collided with professional and procedural norms, which kept their reporting "objective" with less supportive coverage than some had hoped for. As events progressed, substantial economic backlash from investors and advertisers, who felt targeted by protests, turned the table. Suddenly economic and organisational demands became more noticeable to journalists. After initial support, journalists subsequently ceased to portray the protests in a favourable light (see also Shultziner and Shoshan, 2018). This is in line with Curran and Seaton (2007), who have argued that the historical control of the press by political actors has been replaced over time by market forces, which they believe to be even more effective in eliminating radical ideas from public discourse.

On a broader scale, Bennett (1990) has argued that influences like these lead to a pattern of "indexing" of the range of voices and interpretations of events "according to the range of views expressed in mainstream government debate about a given topic" (p. 106). This means that in times of elite consensus, media will limit its discussion to the opinions expressed by the powerful. Only in times of elite dissensus, the voices of the broader public will be heard. Otherwise, alternative information or narratives are assumed to encounter fierce resistance or outright blockage (Bennett et al., 2006). Studies of war-time reporting, in particular, have lent strong empirical evidence to this supposition (Hallin, 1986a; Wolfsfeld, 1997). Generally, it is thought that indexing hurts protesters' ability to explain the causes of their efforts (Andrews and Caren, 2010).

Specifically, reporting in the UK has been described as strongly centred, just as in other countries where free market ideals reign the press (Hallin and Mancini, 2004). Opinions left and right of the Labour and Conservative spectrum are hardly ever supported by the press. Therefore, protest goals outside the spectrum of "normal" political debate might hardly have a chance to gain traction through the press. Curran and Seaton (2007) argue that this is a result of market forces that limit the scope of opinions in the press in Britain. In the 1970s and 1980s, for example, many radical and left-wing outlets were discontinued due to a lack of advertisement and subsequently funding.

Depletion of the market has led to a highly concentrated ownership structure on the national newspaper market: According to a report by the Media Reform Coalition (2021), only three companies (*News UK*, *Daily Mail Group* and *Reach*) control over 90% of national newspaper sales.

Theory-wise, the idea that deviant goals are less likely to be cast in a positive light is hence well supported. However, the conceptualisation of deviance is less clear cut. Interestingly, Boyle et al. (2004) conceptualise the *level of deviance* by using a group's goals as well as the form of protest. The idea is that while street theatre, costumes, civil disobedience or clashes with police are one indicator of deviance, the degree of change sought by a protest is another (also McLeod and Hertog, 1999). This means that a group is deemed deviant either because they do deviant things or because their goal is considered deviant from the mainstream opinion. In later work, however, Boyle and Armstrong (2009) proposed that goals and tactics should be treated as separate factors that exert their own influence on news treatment of protesters.

How extreme or deviant a goal is depends on the assessment of the degree of change sought by the group. The level of deviance of goals is, therefore, often operationalised as three categories: protests that seek to maintain the status quo (least deviant); seek minor reform; or want radical change to the status quo (most deviant) (Boyle and Armstrong, 2009; Boyle et al., 2004, 2005, 2012). The distinction of protesters' goals between support or threat to the status quo, however, is difficult, if not arbitrary, as Kilgo and Harlow (2019) note: there is no irrefutable conceptualisation of the "status quo" as political and social realities are constantly changing; a goal that was a threat to the status quo five or ten years ago could be in favour of the current situation. Boyle and Armstrong (2009), for example, tested this idea by comparing the treatment of pro-life and pro-choice protests before and after the ruling in the Roe v. Wade case — which effectively legalised abortion in the entire US. They found only partial support for the importance of the status quo for protest coverage as pro-life protesters were treated more critically after their goal became to change the status quo, but pro-choice protesters were treated the same before and after the law changed. Similarly, Boyle et al. (2005) found that while the level of deviance of most protest movements decreased over time, headline and article valence stayed roughly the same. Analysing the deviance of protest goals from the status quo, therefore, appears less meaningful.

Besides the level of deviance, studies also assessed the type of goal as a factor for explaining coverage of protests. The specific protest categories differ between studies: Boyle et al. (2004) and Boyle et al. (2005) provided that in a range of events between 1960 and 1999, anti-war protesters were treated most critically, followed by labour protests (e.g., strikes and pickets), protests against wrongdoings by the police (only Boyle et al., 2004) and social-issue protests (e.g., abortion or hunger) which were treated most favourably. In an analysis of newer coverage (2007-2009), Boyle et al. (2012) found that *political protests*, by which they mean protests seeking political change, were treated more critically than anti-war and social-issue protests.

Instead of ranking these types of protest according to the level of deviance, Kilgo and Harlow (2019) suggest that “journalistic routines are subject to a *hierarchy of social struggle*, in which certain topics are given precedence and legitimacy, and others are delegitimised, trivialised, or ignored altogether” (p. 16). They also add several new types to this list — only some of which apply to the UK. Specifically, they found that coverage of protest against anti-Black and anti-Indigenous racism, international protest, and anti-Trump protests use frames that delegitimise their issue far more often than other protest types. For some issues, like health, the environment and immigration protests, the press used legitimising frames more often. Similarly, Harlow et al. (2020) coded 33 specific types of protests before summarising them into six categories. They find that while not all protest types had a significant influence on the use of frames, some patterns emerged: conservative protests, for example, were less likely to be covered in a riot frame but more likely to be covered in a spectacle frame, while socio-economic protests were more likely to be covered using the riot frame. For a legitimising frame, dubbed debate frame, no significant relationship to a protest type was found.

Following Kilgo and Harlow (2019), I expect that some goals are more likely to receive legitimising and some are more likely to receive delegitimising coverage. In other words, I expect that a *hierarchy of social struggle* will emerge in UK reporting of protest, which can predict which causes are given precedence over others. I will use five broad categories for protest goals, ordered from most to least likely to receive critical coverage these are: *political*, *anti-war*, *labour*, *police* and *social-issue* protests (Boyle et al., 2004, 2005, 2012). The hypotheses I will test are thus:

- **H3a**: The goal of a protest determines how likely delegitimising framing is to be used, following a hierarchy of social struggle that is ordered from highest to lowest probability: *political*, *war*, *worker*, *police* and *social issues* protests.
- **H3b**: The goal of a protest determines how likely legitimising framing is to be used, following a hierarchy of social struggle that is ordered from highest to lowest probability: *social-issue*, *police*, *labour*, *anti-war* and *political* protests.

Note that the expected hierarchy in **H3b** is the flipped version of **H3a** as I expect that the likelihood for a protest to be covered with a legitimising frame should be roughly diametrically opposed to the likelihood for a protest to be covered with a delegitimising frame.

Tactics. As mentioned above, early studies often treated goals and tactics as two sides of the same coin. However, Boyle and Armstrong (2009) established that this is a shortcoming as both have distinct effects on the coverage of protest. Specifically, they showed that when both are taken into account, extreme protest tactics facilitate critical coverage stronger than protest goals. This insight had already been observed by one of the earliest studies of protest coverage: Gitlin (1980) found that the amount of outright negative coverage, instead of trivialisation, increased as anti-war protests became more extreme. Boyle et al. (2012), who additionally test the effects of protest location, protest type, and a group's goals in an international comparison of protest coverage, conclude that the tactics employed during a protest are the strongest determinant of delegitimising coverage. Likewise, Wasow (2020) concludes that nonviolent black-led protests between 1960 and 1972 led to more coverage of civil rights issues, while protester violence invoked language associated with disorder and calls for law and order.

How extreme the tactics are is usually determined by two variables: do the protesters break laws and do they engage in violence? From a broader perspective, a focus on extreme protest tactics is, therefore, in line with the professional norms of "newsworthiness" mentioned above: usually, the clashes rather than the message of a protest will produce most conflict, drama, unusualness and negativity. This leads to the following expectations:

- **H3c**: When protesters break laws or when they engage in violence, delegitimising framing is used more often.
- **H3d**: When protesters are peaceful and obey the law, legitimising framing is used more often.

The relation between violent tactics and coverage has been known for a long time by both researchers and protesters (Gitlin, 1980). When tactics become more violent, coverage becomes less supportive of the cause and more supportive of repressive measures. Nevertheless, protesters often use violent tactics because this will usually lead to (more) coverage (Boyle et al., 2012). **H3c** and **H3d**, therefore, also answer the question if this bargain makes sense for a movement.

State Response. As mentioned above, Wasow (2020) found in his study of black-led protests between 1960 and 1972 that protester-initiated violence led to delegitimising coverage and subsequently turned public opinion and voter preferences against the civil rights movement. However, Wasow (2020) also tested the effects of violent state response towards the movement. Specifically, he found that there are four scenarios that yielded slightly different outcomes: When the state tolerated peaceful protest, this lead to coverage which legitimised the concerns voiced by protesters; when the state used violent repression against peaceful protesters, this lead to coverage that is more supportive of the protesters' goal — and to an overall higher volume of coverage about the respective event; when only the protesters were violent, the civil rights movement was penalised in the coverage; the same was, however, true when both police and protesters engaged in substantial violence. In this last case, the coverage is predicted to focus on concerns about order and public safety.

Wasow's (2020) study thus mostly confirms the relationship between violence and (de)legitimising coverage formalised in **H3c** and **H3d**. A scenario in which the state uses violent repression against peaceful protesters, however, has not been covered by any previous studies. Following Wasow (2020), I include this additional layer, which leads to the expectation that:

- **H3e**: When the state uses repressive tactics on peaceful protesters, legitimising framing is used more often.

3.4.2 Outlet-Level Factors

Newspaper Ideology. In the landmark study which defined the idea of a *protest paradigm*, Chan and Lee (1984) found that not all newspapers engage in delegitimising protests in the same way: the left-leaning media were less hostile towards the protesting teachers and students in Hong Kong than centrist media, which were in turn less hostile than right-leaning journalists. This factor has usually been overlooked in subsequent US-based studies, which might assume that ideological differences between major outlets are less important given the emphasis on objectivity in the country's journalistic culture. More recent studies, however, have started again to employ newspaper ideology as an explaining factor for the level of marginalisation different protests face (Lee, 2014; Oz, 2016; Shahin et al., 2016). As mentioned above, UK media are distinctly more partisan than outlets in the US and elsewhere (Hallin and Mancini, 2004; Kuhn, 2007), which is likely to increase the effect of newspaper ideology on protest reporting. This leads to the expectation:

- **H3f**: Right media outlets use delegitimising framing more often than left media outlets.
- **H3g**: Left media outlets use legitimising framing more often than right media outlets.

Weaver and Scacco (2012) add nuance to this: more than by their right or left ideology, the coverage of an outlet might be determined by the distance of their own ideological tilt to the ideology of a protest in question. They test this using prime-time cable news coverage about the right-leaning *Tea Party* movement in the United States in 2009-2010 to assess which outlets used marginalisation devices more often. Their finding suggests that liberally aligned MSNBC was most likely to marginalise the *Tea Party*, followed by the centrist AP and CNN, and Fox News being supportive overall. Since protests have traditionally held goals that were considered left more often in Western countries — and continue to do so (Borbáth and Gessler, 2020) — the assumption that right-leaning media are generally more critical of protest might be incorrect. This leads to two hypotheses that will be tested:

- **H3h**: When there is a divide between the ideology of the protest's goal and the ideological leaning of the outlet, delegitimising framing is used more often.
- **H3i**: When there is agreement between the ideology of the protest's goal and the ideological leaning of the outlet, legitimising framing is used more often.

Type of Newspaper. Gavin (2007, pp. 95-118) and Rosie and Gorringe (2009) add another factor related to newspaper outlets that is especially pronounced in the UK: the difference between tabloid and broadsheet newspapers. Gavin (2007), who assessed the May Day demonstration in London in 2000, concluded that tabloid outlets used elements from the *protest paradigm* more extensively and failed to mention any motivation or goals of the protest. Broadsheet newspapers — especially *The Guardian* — apparently made an effort to explain why the protests were held. In their qualitative study, Rosie and Gorringe (2009) find that tabloid newspapers were more extreme in their framing of peaceful G8 protesters as a fringe group of radical militants. The sharp divide in reporting style between broadsheets newspapers (*The Daily Telegraph*, the *Financial Times*, *The Guardian*, *The Independent* and *The Times*), and tabloids (*Daily Express*, *Daily Mail*, the *Daily Mirror* and *The Sun*) is well-known and essentially unique to the UK media system. In general, it is expected that tabloid newspapers are more focused on human interest and entertainment (Hallin and Mancini, 2004; Kuhn, 2007). It is, therefore, only plausible to expect that the two types use different frames. Following the work by Rosie and Gorringe (2009), the following expectations can be made:

- **H3j**: Tabloid media outlets use delegitimising framing more often than Broadsheet outlets.
- **H3k**: Broadsheet media outlets use legitimising framing more often than Tabloid media outlets.

3.4.3 Time-Bound Factors

Two factors related to time are expected to have an influence on reporting about protests. The first one was already mentioned above: the year in which an article was published is expected to determine if it is being framed in a delegitimising way, with newer articles using delegitimising framing less and legitimising framing more often (**H2a** and **H2b**).

The second factor very likely only affects large scale, highly mediatised protests. Gottlieb (2015) suggests that coverage follows a *news framing cycle*: coverage of protest begins with a focus on conflict — hence marginalising or ignoring the goals and grievances of a protest — then shifts to more substantial coverage before returning to conflict. This concept, based on Downs (1972), follows the logic that journalists will use different frames to cover an event as time passes "to keep the story alive and fresh" (Chyi and McCombs, 2004, p. 22).

The *news framing cycle* intuitively makes sense, yet the proposed time frame might suffer from one problem: Gottlieb's (2015) case, the Occupy Wall Street movement, was arguably special since protest stretched over a long period and media attention did not wane for months. He, therefore, suggests that the second phase, when reporting shifts to more substantial coverage, started only in the third week after the protest started. This finding is probably not suitable for generalisation as attention to most protests wanes after days, rather than weeks or months. In most cases, the media will have forgotten about a protest the day after the first reports were printed.

Nevertheless, the idea of a protest news framing cycle seems worth testing, albeit in an abbreviated form:

- **H3l**: At the beginning of protest coverage of an event, news will be event-driven and hence are more likely to contain delegitimising framing.
- **H3m**: The more time passes between the start of a protest event and publication of an article about it, the more likely it is that the article contains legitimising coverage.

Chapter Conclusion

This chapter has introduced the theoretical framework for this thesis. It laid out the expectations for how the news media cover protest in the United Kingdom. Specifically, Table 3.1 summarises the hypotheses posed in this chapter.

The first part of the chapter argued that *framing* is the most suitable theoretical lens through which to examine protest coverage for two reasons: it helps to systematise the study of content as frames are recognised as the guiding principle for how content is created. Studying frames can thus help to condense heterogeneous media content into a relatively small number of clearly defined categories. Furthermore, I argued that the framing literature also provides a better theoretical framework to guide the interpretation of emerging reporting patterns. While framing effects and public opinion do not play an active role in this thesis, knowledge provided by the literature about it can help to draw conclusions on what the salience of specific frames actually means. Overall, I follow the expectation from the *protest paradigm* literature that poses that protest coverage will be primarily framed delegitimising (*H1*).

The second part of the chapter has then argued that we can not assume that media coverage of protest has stayed unaltered since the seminal studies in the field about movements in the 1960s. Specifically, developments in the media market, journalists' sense of consumer demands, additional gatekeepers in the emerging "hybrid media system", and the broader society are arguably very different today than in previous decades. Specifically, I expect delegitimising framing to decrease and legitimising framing to increase in salience over time (*H2a–b*).

The final part of the chapter surveyed what factors have been argued to influence the qualities of coverage different protests receive. I argued that on the event level, the goals and tactics of a protest and how the state responds to protest events leads to different kinds of reporting (*H3a–e*). On the outlet level, the ideology and type of newspaper covering a protest is expected to shape whether an event receives legitimising or delegitimising coverage (*H3f–k*). Finally, I pose that time has an influence on reporting. As said in the second part of the chapter, the year of publication will shape media content, as ongoing trends in the media landscape and society are expected to matter for reporting. However, time is also expected to play a role as the reporting about a specific protest is expected to change depending on how much time passes between a protest and the publication of a report about it (*H3l–m*).

Table 3.1: Hypotheses

	Hypothesis
H1	Stories on protest events mainly use delegitimising framing as described in the protest paradigm literature.
H2a	Delegitimising framing decreases in salience over time.
H2b	Legitimising framing increases in salience over time.
H3a	The goal of a protest determines how likely delegitimising framing is to be used, following a hierarchy of social struggle that is ordered from highest to lowest probability: *political*, *war*, *worker*, *police* and *social issues* protests.
H3b	The goal of a protest determines how likely legitimising framing is to be used, following a hierarchy of social struggle that is ordered from highest to lowest probability: *social-issue*, *police*, *labour*, *anti-war* and *political* protests.
H3c	When protesters break laws or when they engage in violence, delegitimising framing is used more often.
H3d	When protesters are peaceful and obey the law, legitimising framing is used more often.
H3e	When the state uses repressive tactics on peaceful protesters, legitimising framing is used more often.
H3f	Right media outlets use delegitimising framing more often than left media outlets.
H3g	Left media outlets use legitimising framing more often than right media outlets.
H3h	When there is a divide between the ideology of the protest's goal and the ideological leaning of the outlet, delegitimising framing is used more often.
H3i	When there is agreement between the ideology of the protest's goal and the ideological leaning of the outlet, legitimising framing is used more often.
H3j	Tabloid media outlets use delegitimising framing more often than Broadsheet outlets.
H3k	Broadsheet media outlets use legitimising framing more often than Tabloid media outlets.
H3l	At the beginning of protest coverage of an event, news will be event-driven and hence are more likely to contain delegitimising framing.
H3m	The more time passes between the start of a protest event and publication of an article about it, the more likely it is that the article contains legitimising coverage.

Chapter 4

Research Design and Methods

This chapter outlines the research design, data, variables and methods used to assess news media coverage of protests considering the hypotheses developed above and listed in Table 3.1. In short, the design involves a single country case study analysing content in eight major national UK newspapers. The aim is to assess coverage of protest and explain varying patterns with time-bound, event and outlet level factors. The study covers content on a diverse range of protest events in outlets divided by type and ideology during a relatively long time period of 26 years (1992-2017). This ensures a high variation of observations and maximises the scope of the study.

The chapter is divided into three parts. First, Section 4.1 outlines the fundamental decisions regarding the research design: the choice of the United Kingdom, the specific time frame and the scope of the analysis. These decisions are discussed in light of how this thesis aims to overcome the limitations in the existing literature identified in the review in Chapter 2. The ultimate goal is to provide broader, more systematic and longitudinal evidence to advance our understanding of how protests are covered by mainstream news media.

To do so, the research strategy follows a two-step procedure. Section 4.2 discusses the first step: how the dependent variable — media frames in protest reporting — is conceptualised and measured. This step is not only necessary to provide the data for further analysis, it also constitutes the main methodological contribution of this thesis. In short, the approach combines dimension reduction with a supervised learning approach to perform a large-scale assessment of media frames. This makes it possible to systematically examine patterns of protest reporting on a scale that is, to my knowledge, unprecedented.

The third and final part of this chapter covers the independent variables for the second analysis step: scrutinising in which ways the identified factors play a role in shaping the framing of protest in UK newspapers. Specifically, I discuss the employed data and explain the strategy to match the dataset on media coverage of protest with the dataset on protest events employed in this thesis.

4.1 Case Selection

This thesis analyses coverage of domestic protest in mainstream newspapers, based in the United Kingdom, from 1992 until the end of 2017. The specific case was chosen in an attempt to fill several gaps in our knowledge about how news media cover protest. Furthermore, the case offers unique opportunities to test the hypotheses developed in Chapter 3. As discussed in Chapter 2, there are currently three main limitations to our knowledge about news covering protests: a lack of theoretical clarity and a shared explicit operationalisation strategy regarding the so-called *protest paradigm*, which limits the degree to which studies can be compared through time and cases; most studies cover limited time frames, which means we know relatively little about changes over time, especially as seminal research about the *protest paradigm*-concept is rather dated; and the often rather limited scope of research, which limits our knowledge to specific protest events and issues rather than a systematic knowledge about all covered protests. A strategy to increase theoretical and conceptual clarity was discussed in Section 3.1. In short, this study employs theoretical and methodological advances from framing research to improve the theoretical foundation of the analysis and make the operationalisation of key concepts more explicit, and hence transparent and potentially replicable.

The choice for the United Kingdom for this single-country case study is made for several reasons. First, a single-country design can increase internal validity compared to cross-national designs as unobserved characteristics of the specific media system are held constant (Marczyk et al., 2005, pp. 158–197). In a study comparing multiple countries, these unobserved differences might lead to spurious correlations between the selected variables. This problem is particularly acute for the object of study here as there are numerous differences in media systems as well as protest behaviour, police tactics and the level of acceptance towards protest as a tool for political expression. Controlling for all these differences would be difficult if not impossible — not least due to the knowledge gap identified in Section 2.2.2 about what the differences between countries are in these areas.

Second, specific settings of the UK media economy allow some very interesting insights regarding the outlet-level factors described in Section 3.4. While an ideal of objectivity is traditionally strong in many media systems, national UK newspapers are more open in their partisanship towards one of the main political parties (Hallin and Mancini, 2004). Since the ideological stance of outlets in the UK is generally known, hypotheses about the influence of ideological alignment between a protest's goals and the media outlets can be more easily tested — specifically, vis-à-vis the US media landscape. Additionally, the British newspaper market is rather unique in its sharp segmentation between broadsheet newspapers and tabloids. Since the different outlets target different audiences, the tactics for competition also differ: tabloids try to win over larger audiences while broadsheets focus on a readership in a specific, upper and middle class, socio-economic composition (Kuhn, 2007). This can result in the marginalisation of politics in favour of "human interest" and lifestyle stories in tabloids (Curran and Seaton, 2007; Street, 2011, 88ff), while it is also attributed to creating incentives for broadsheets to provide more analysis and commentary in addition to news reportage

(Kuhn, 2007). It is therefore expected — and has, in fact, be shown before (Gavin, 2007) — that the segmentation between broadsheets and tabloids creates distinctly different characteristics of coverage when it comes to protests. The case of the UK is thus exceptionally well suited to test a number of different factors associated with reporting about protest.

Finally, the UK is especially relevant, since it was one of the countries in which the phenomenon of marginalising coverage of protest groups was first observed (Glasgow University Media Group, 1985; Halloran et al., 1970; Murdock, 1973). This allows for a direct comparison of this research to seminal work of the past.

As has been mentioned above, most previous studies had a rather limited scope. They focused on specific protest movements, on single, often radical, protest events or on a series of events all challenging the same specific issue or policy. Other studies narrowed their population of interest down by using very specific keywords (Boyle et al., 2012) or are restricted to few locations (e.g., Wisconsin in the case of Boyle et al., 2004; and Brussels for Wouters, 2015b). This arguably reduces the generalisability of findings as it is expected that specific features of protests lead to different portrayals of the events in mainstream news media. To increase external validity while answering the research question on the main frames of protest reporting, this research includes all stories, in the selected outlets, about domestic protest events anywhere in the United Kingdom. The long time period that was chosen also enables this research to cover a variety of different events which vary greatly in terms of the chosen independent variables outlined below.

A final issue identified in Chapter 2 is that most previous studies of protest coverage only cover time periods of a few months, weeks or just days of reporting.[7] This means they are basically snapshots of the practices of reporting at a specific point in time. This could limit the validity of the research, considering that media landscapes

[7] Although there are some notable exceptions, especially Boyle et al. (2004), Boyle and Armstrong (2009), Di Cicco (2010), Hughes and Mellado (2015) and Jha (2007).

are changing rapidly, journalistic practices are changing substantially, and norms in society are changing gradually but significantly, as discussed in Section 3.3. Among the important societal and technological changes, new forms of news and social media brought on by the internet as a mass phenomenon probably has had the biggest impact.

The choice for the time frame of this study was, therefore, made with two considerations in mind: First, the long time period spanning 26 years enables this thesis to cover vastly different periods of protest reporting. During the time frame there have been periods of high frequency of protest, high salience of specific movements and relatively calm phases; different governments; as well as a variety of ideologically and organisationally diverse events. Second, to capture the developments around the emergence of the internet, 1992 was determined as a starting point. This year marks the beginning of profound changes in the gathering and distribution of news. At the beginning of the 1990s, the first online newspapers began their work while many of the traditional outlets started to make some of their stories available online (Stanyer, 2010). Furthermore, 1992 also was the first year in which a protest group in the UK started to use the internet as a means of communication (McKay, 1998). Both of these trends were, however, merely precursors of the mass phenomena that set in at the start of this century. The chosen time span thus allows this thesis to study the effects of both trends roughly from their start, up to today.

4.2 Dependent Variable: Reporting of Protest

I use media coverage of protests as the dependent variable in this thesis. Specifically, I measure the framing used in newspaper reports on protests and demonstrations and construct one binary variable per frame — which takes the value 1 if a frame is present and 0 otherwise. The decision to code newspaper articles instead of alternatives, such as online news, social media content or TV transcripts was based on two arguments. Despite declining audiences, mainstream news media such as newspapers are still deemed the most important platforms for public debate and the creation of the public agenda

(Rogstad, 2016). Newspapers might have lost some of their intermedia agenda-setting power due to being slower to release breaking news, but the more in-depth writing produced by newspapers nevertheless forms a crucial share of what is redistributed via online news and social media and is key in initiating, amplifying and sustaining attention to an issue (Harder et al., 2017; Langer and Gruber, 2021; Rogstad, 2016). Furthermore, suitable data from UK newspapers were already archived and digitally available for the whole period of interest.[8] This is especially important since the ideology and type of outlets are thought to matter for the type of reporting, which would lead to problems if outlets changed during the studied period.

I use a dataset specifically created for this thesis, which was compiled by using a keyword search on the newspaper database of *LexisNexis*. This dataset was subsequently filtered in several steps and with several different techniques to only contain reports about protests. Details of the dataset creation are discussed in Chapter 5. This data is used to measure the dependent variable using content analysis as explained below. Overall framing patterns will be scrutinised in Chapter 6 before I use multilevel logistic regression models for each frame to explain emerging patterns in Chapter 7.

This section is divided into three parts: the first defines which kinds of events are treated as protest in this thesis. The second one surveys available approaches to frame measurement and automated content analysis. The third introduces a dedicated new method to identify and code frames in media reports about protest on a large scale.

4.2.1 Definition of Protest

To get a clear understanding of what kinds of events *protest* comprises, the term first needs to be defined. However, this is not a trivial task as there is considerable debate over what counts as *protest*. Opp (2015), who compared several definitions of protest from the social movement literature, found that definitions commonly share four aspects: (1) protest is commonly perceived as a joint action, as opposed to the dissent of

[8] Although not for all newspapers, as described below.

individual citizens; (2) actors within a protest usually share at least one goal which is typically to express one or multiple grievances; (3) the actors cannot directly achieve a goal themselves but want to influence either the public or official decisionmakers — this presupposes protesters express their objections publicly (also see Lipsky, 1968; Turner, 1969); and (4) that protest behaviour is irregular as opposed to e.g. party conventions, meetings of a parliament or elections which follow institutional rules.

The first three of these features seem relevant and fit closely with the events scrutinised in the available literature on the *protest paradigm* — although explicit definitions of protest are generally not provided. The last aspect — irregularity of events — is understood here as non-institutional, given that recurring events, such as May Day protests, do seem worthwhile to be included, even though they are repeated each year.

Protest is therefore operationalised in this thesis as: *a collective overt public expression that either articulates grievances against or support for one or multiple targets (i.e., policies, institutions or behaviours) in order to either directly influence an institution's decisionmakers or the knowledge, attitudes, and behaviours of the public.*

This includes, but does not limit the analysis to, social protest. Defined as "a form of political expression that seeks to bring about social or political change […]" (McLeod, 2011), the term social protest limits the number of cases to those protests that seek change whereas, for example, counter-protesters often explicitly demand to preserve the status quo. An alternative would have been to only include political protest, yet, while protests which target policymakers were found to make up the large majority of cases in the UK, protests which engage in public criticism of a company's or a person's behaviour were not excluded. Furthermore, protest is understood here as a resource that can be used, for instance, by social movements. Nevertheless, if the group which initiates a protest is a social movement or not, does not play a role in defining which events and actions are counted as protest. A protest is instead initiated by a *protest group*, which is understood as any collective of actors who engage in protest as defined here.

Additionally, two important practical restrictions were applied: First, only domestic protests were taken into account for theoretical clarity. Research has shown that non-domestic protest receives significantly different reporting compared to domestic protest events (Boyle et al., 2012; Mueller, 1997). This is because foreign protests usually do not target actors within the same country the outlet describing it operates. This means that most restraints journalists are thought to face when reporting about protests are absent in such a case. A second reason is that protests themselves operate within the political and economic system (Lipsky, 1968). Especially protests in non-democratic countries, such as the recent examples of the Arab Spring or the Maidan insurrection, thus operate according to a different logic compared to protests in the United Kingdom. The focus of this thesis lies on protest as a democratic resource, not on protest as a mean to overthrow a despotic government, which means that further questions are out of scope for the moment.

The second restriction is that articles about sectarian violence in Northern Ireland were excluded. It could be argued that at least some of the events in this category could fall under the definition of protest presented above. In the *Drumcree Standoff*, for example, the attempts by Orangemen to organise a parade, as well as actions of those opposing a parade through a mainly Catholic/Irish nationalist part of town include overt public expressions that articulate grievances. Similar to non-domestic protest, though, reporting about these actions is structurally different. The high-profile conflict between Unionists and Irish nationalists was complicated and lasted for decades. Thus journalists could not and/or did not want to recount the conflict lines, which the public presumably knew about anyway, for every event in the 1990s and 2000s. During coding of the articles, it emerged that reports of sectarian violence in Northern Ireland did not feature any explanation of goals and grievances, only focused on violence and usually described events as riots rather than protests. Like non-domestic protest, sectarian protests in Northern Ireland are an interesting case to be researched, yet both aspects lie outside the scope of this thesis.

4.2.2 Measurement of Media Frames — Available Approaches

As with other areas in framing research, there is no shortage of different approaches to measure media frames. Consequently, *how* frames can be extracted reliably from text in order to be analysed is still disputed. In essence, studies try to find patterns in how aspects of an issue or story are selected and made salient in media content and describe those in terms of different frames. However, meta-studies such as Matthes (2009) or Entman et al. (2009) reveal a plethora of different approaches on how frames should be operationalised and measured. Studies have been using text-based and number-based approaches, have studied frames inductively or deductively and have extracted generic or specific frames from the material. To situate the approach employed here in the field, I discuss four categories of approaches: *qualitative* approaches, *manual-holistic* approaches, *semi-automated* and *fully automated dimension reduction* approaches.[9]

Qualitative approaches identify frames based on the interpretation of the text itself. Usually, these approaches are rooted in qualitative research traditions, proceed inductively, frames are described in-depth and little or no quantification of elements or the distribution of frames within a discourse is provided by the researcher. Since coding is complex, it is also more labour intensive than other approaches, which is likely why most of the studies in this category rely on small samples of text which renders generalisation difficult. Notably, defining studies in *protest paradigm* research, especially McLeod and Hertog (1999) and McLeod and Hertog (1992), fall in this category. Like with many examples of this category, they offer a thorough description of the individual frames. Yet, studies in this category usually do not offer the same level of detail when it comes to questions about how researchers arrive at their conclusions. How did this specific number of frames emerge from the material and which are their distinctive codable features? If these questions can't be answered, replication is difficult. Additionally, a researcher employing an inductive qualitative approach runs an increased

[9] The distinction is based on more detailed overviews by Entman et al. (2009), Matthes (2009) and Matthes and Kohring (2008) who arrive at similar categories.

risk of identifying frames “they are consciously or unconsciously looking for” (Matthes and Kohring, 2008, p. 259).

Secondly, *manual-holistic approaches* code frames as holistic variables usually in quantitative content analyses. *Content analysis*, which is also employed in this study, is defined as “a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use” (Krippendorff, 2004, p. 18). Essentially, researchers look for patterns and themes in recorded communicated human messages or their context, which makes them more accessible for summary and quantitative analysis. Traditionally content analysis research is done by a team of coders and results are only deemed reliable if the agreement between coders is found to be sufficient (Krippendorff, 2004). As in qualitative approaches, the validity and reliability of an approach depend on the transparency with which the study communicates the coding decisions.

Two issues are common regarding *manual-holistic approaches*. First, even if the translation of frame descriptions into content analytical codes, which is difficult itself, is done commendably, coding frames in content analysis remains decidedly challenging (van Gorp, 2007). Frames are an abstract variable and notoriously hard to code in content analysis as the decision if a frame is present or not involves a non-trivial degree of inference (van Gorp, 2005). Due to the immense complexity of frames, content analysis which uses holistic frames as content analytical variables is, therefore, prone to receive a subjective tint based on the researcher’s perception of the material (Matthes and Kohring, 2008). If coding decisions follow transparent criteria and if inter-coder reliability can be established, this issue can, however, be mitigated.

A second issue might be more substantial: while *manual-holistic approaches* can work relatively well in the frame *coding* phase of research, they are less formalised and more prone to reliability and validity issues in the frame *identification* phase. Essentially, in approaches in this category, frames can be either derived from the literature or identified inductively in a pilot study of a small sample. Derived from the literature usually means that studies are based on categories from previous qualitative research. Frames based on a pilot study are ideally found by multiple researchers who arrive at similar categories. In both cases though, the exact processes behind the identification of frames remain a black box. Furthermore, once a list of frames is defined, new frames can usually not emerge from the material. That means that if a relatively uncommon frame is found only later in the coding process, researchers will usually not change their initial assessment or will not notice a new pattern at all (Matthes, 2009).

To evade problems of reliability and validity, researchers have turned to approaches of *semi-automated* and *fully automated dimension reduction*. Early on, studies have tried to improve coding reliability and reduce manual labour by using *fully automated* approaches and thus replacing manual coding completely. Miller (1997), for example, proposes so-called "frame mapping", which employs a combination of cluster analysis and multidimensional scaling based on the co-occurrence of key terms in texts. The key terms are chosen based on frequency and researcher assessment of relevance to the topic. The results must be validated by a human. Otherwise, no manual coding needs to be done. This reduces the subjectivity of frame *identification* significantly, as the clusters are extracted by the computer via statistical methods instead of being found through interpretation by the researcher.

Instead of selecting a set of words manually, more recent *fully automated* approaches employ the complete set of vocabulary or employ pre-defined rules to exclude only words that are thought to be irrelevant.[10] In a nutshell, these dimension reduction techniques of text analysis assess the vocabulary of text as dimensions. Each text has a value of one or more on a dimension if it employs a word once or multiple times, or it has a value of 0 on this dimension if a word is not employed. Techniques exist which meaningfully reduce the dimensionality of such text data by reducing the thousands or hundreds of thousands of dimensions into a human interpretable set of

[10] Such as stopwords, which are unlikely to convey much information (e.g., "the", "was", "to"), very frequent words, which occur in almost all documents and rare words which are usually irrelevant for category formation anyway.

latent dimensions. These latent dimensions are thought to reflect the same information given by all words, yet reduced on a small set of meaningful categories. Currently the most advanced of these approaches is called *topic modelling* and is usually based on Latent Dirichlet Allocation (LDA) (Blei et al., 2003). Effectively, LDA extracts the main themes from even massive collections of documents. These so-called topics are usually more coherent and semantically interpretable than groups based on other technique, such as cluster or factor analysis (Blei, 2012; Nicholls and Culpepper, 2020).

Several studies have used topic models to operationalise frames (e.g., Baumgartner et al., 2008; DiMaggio et al., 2013; Gilardi et al., 2021). However, operationalising frames as topics in a topic model has two important limitations: first, while topic models take word choices into account — which are an important part of framing (Entman, 1993) — other choices, regarding what information is presented in a text and what information is left out, do not play a role. Frames extracted by these methods have, therefore, been argued to not do the concept full justice as they reduce frames to clusters of words that are used together while omitting other dimensions (also see Carragee and Roefs, 2004; Matthes and Kohring, 2008).

Besides this conceptual argument, there is also an important empirical consideration. Jacobi et al. (2015) used LDA on a large media corpus containing coverage of nuclear technology in the US spanning several decades. Theoretically, the extracted topics should be congruent or comprise the frames identified in the pioneering framing analysis study by Gamson and Modigliani (1989) who used a similar dataset. However, while the original study identified several "anti-nuclear" frames, no topic could be clearly linked to that frame — even though many reports about nuclear accidents and the dangers of nuclear power employed this frame.

This is because LDA extracts the main topics or themes in a corpus of text based on its words. These themes might or might not coincide with frames. In this thesis, LDA was used to clean the dataset of news media articles (see Chapter 5). Instead of framing, the emerging topics signalled the different kinds of protest: anti-war, anti-nuclear, student protests and so on. Topic models appear more suitable for framing analysis in homogeneous corpora in which all documents focus on the same or very similar stories and only differ from each other in the choice of vocabulary. In broader corpora, such as the dataset on protest coverage in UK newspapers analysed here, topic models will pick up broader themes while generally ignoring distinctions between frames. This is essentially the conclusion of a recent study by Nicholls and Culpepper (2020) who compare different dimension reduction techniques, namely k-means clustering, evolutionary factor analysis and structural topic models, with manually identified frames: the topic model approach was the only one to succeed in finding categories similar to the human identified frames in a highly homogeneous corpus. Yet it still failed to do so in a broader dataset.

Finally, there are *semi-automated dimension reduction* approaches, which are employed in this thesis. They are based on the idea to manually pre-select the dimensions used as a starting point in dimension reduction. In practice, this is done by splitting up frames into sub-variables which are easier to code in manual content analysis. The results can then be fed to the same dimension reduction techniques mentioned above. The resulting latent dimensions tend to be closer to human identified frames compared to fully automated approaches, as researchers pre-select meaningful features of a text before the automated analysis step. Furthermore, semi-automated dimension reduction outperforms human capabilities in identifying frames in a valid and reliable way as the individual coding decisions are smaller, easier to make and more transparent. In some regards, the procedure resembles approaches that analyse survey questionnaires in which subjects are asked a number of questions from which researchers draw conclusions about deep-seated characteristics and beliefs.

Consequently, frames are often operationalised as a set of indicator questions. Coders can then "interview" documents, assessing whether a certain aspect is mentioned in the text or not. This breaks up coding into smaller, more clearly defined categories, making coding decisions more manageable, more transparent and thereby improving the reliability and validity of framing analysis considerably (Matthes and Kohring, 2008; van Gorp, 2007). However, the way the dimension reduction of indicator questions was previously approached is rather simplistic: indicator questions are usually structured around the idea that questions belong to pre-defined frames and that a frame is present in a text if any of the respective questions are answered with “yes” (see e.g., Burscher et al., 2014; Card et al., 2015; de Vreese et al., 2001; Semetko and Valkenburg, 2000). The downside of this approach then is that frames must be known — and well-understood — ex-ante, since they form the base of the indicator questions. A solution is to either draw frames from the literature or to conduct a pilot, similar to the ones conducted by manual-holistic approaches. Dimension reduction approaches which use indicator questions can thus be seen as an improvement to *manual-holistic approaches*, yet still suffer from some of the same problems in the frame *identification* step.

Matthes and Kohring (2008) suggest that instead of indicator questions, the *frame elements* included in the definition by Entman (1993), cited in Section 3.1, are an ideal starting point for meaningful dimensions. Specifically, Entman (1993) suggests that a frame consists of a *problem definition*, *causal interpretation*, *moral evaluation*, and *treatment recommendation*. These elements can be further divided into content analytical variables which can be clearly defined and coded relatively easily. In this understanding of framing, a frame is a pattern of different frame elements used together in a text. The procedure then identifies common frames using dimension reduction techniques on the coded elements.

This improves the reliability and validity of the identification of frames substantially. The reasoning is essentially the same as for coding frames with indicator questions: instead of the complex decisions involved in identifying the main organising framing patterns in the discussion of an issue, researchers can identify smaller, clearly defined units. For example, who are the most important actors? Which topics are repeatedly discussed? Which problems are identified? As frames are thought to be latent structures underlying a story (van Gorp, 2005), established methods to uncover latent structures can also be employed. Theoretically, these methods can perform both steps in content analysis of framing: the inductive step of frame *identification* and the usually deductively performed step of frame *coding* can be approached using the same coded data and methods. Techniques such as k-means clustering and factor analysis provide information about latent dimension and membership of units to classes (single-membership approaches) or prevalence of classes in units (mixed membership approaches).

Finding the main frames in a discussion inductively, however, does not mean that previous work must be ignored. In this thesis, the marginalisation devices embedded in the *protest paradigm* serve as a starting point from which I develop the codebook for this study. The researcher can split frames known from the literature into elements, while a new *problem definition*, for example, can be added to an existing coding scheme during a pilot phase or later. Therefore, the approach by Matthes and Kohring (2008) appears as the most promising starting point to conduct valid and reliable framing analyses.

While dimension reduction techniques can perform frame *coding*, the approach comes with an important caveat: they need manually coded data first. As mentioned, researchers tried early on to automate frame *coding* procedures as manual coding is expensive — time and resource-wise — and costs quickly add up as the scale of research expands. This also poses a problem for this thesis. Analysing the entire population of interest via manual content analysis would not be feasible, given the broad scope: all stories about protest in mainstream news media over the chosen 26-year period. One option, in this case, is sampling. However, as mentioned above, artificially decreasing the scope would be detrimental to the project while random sampling would decrease the efficiency of the analysis and introduce uncertainty (King et al., 1995, pp. 66-74).

This is where Automated Content Analysis (ACA) methods — sometimes called quantitative text analysis or text-as-data methods — come into play as they promise to radically decrease the costs of analysing content (Grimmer and Stewart, 2013). The remaining sub-section thus reviews existing ACA approaches in regard to their suitability for this thesis. ACA methods, originally developed by computer scientists, are used more widely and commonly among social scientists today as indicated by reviews in political science (Grimmer and Stewart, 2013), Journalism (Boumans and Trilling, 2015), Sociology (Molina and Garip, 2019; Nelson et al., 2021) and Political Communication (Welbers et al., 2017), since they offer the chance to make large-scale content analysis feasible without an overwhelming amount of resources.

As described, *fully automated methods* like LDA were deemed not suitable for the project. However, ACA comprises many different procedures. Specifically, Boumans and Trilling (2015) classify ACA techniques along a continuum ranging from inductive to deductive approaches. *Fully automated dimension reduction* techniques, like topic models, are inductive as they analyse content with very few or completely without any assumptions of what is meaningful in a corpus of texts. In other words, the discovery of patterns is entirely left to the computer (Boumans and Trilling, 2015).

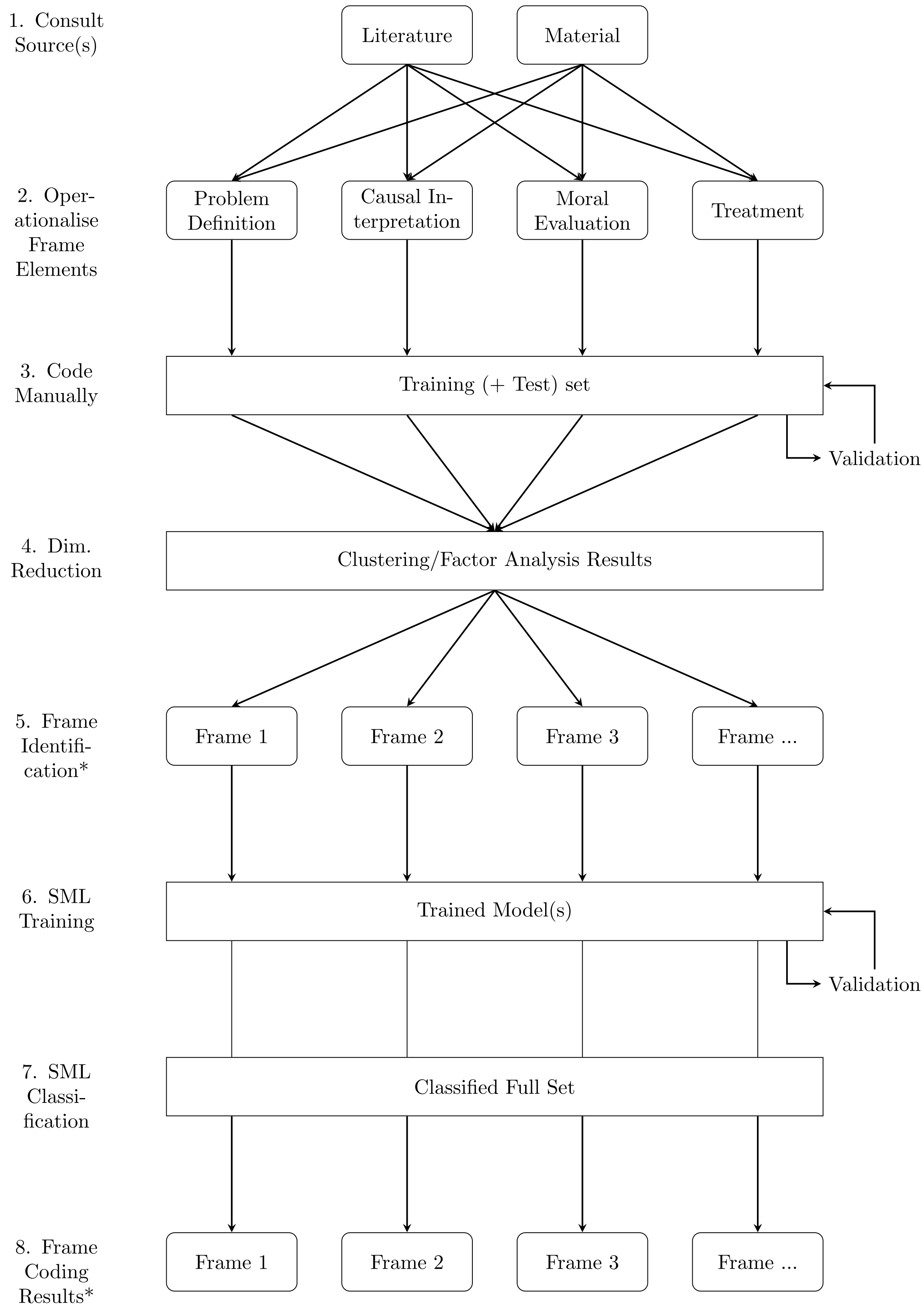
On the other end of the spectrum are deductive methods which assume that the researcher knows exactly what s/he is searching for and only the coding or classification is automated. The most deductive approaches are counting- or dictionary-techniques: researchers construct a list of meaningful tokens — these can be single words, sequences of words or search patterns — which best indicate the differences between texts in previously identified meaningful categories. One of the most widely employed applications for these kinds of techniques is sentiment analysis. Most commonly, researchers rely on a pre-defined dictionary that comprises positive, negative and sometimes neutral words. A text which consists of more positive than negative words is deemed to be in the positive category — at least in the simplest form of this approach (Young and Soroka, 2012). For this thesis, an approach using dictionary methods would consist of first identifying the main frames, then constructing a suitable dictionary from the respective categories of text before applying the dictionary to the corpus. However, constructing a dictionary is very labour-intensive; the resulting dictionaries are highly specific to the domain or even corpus they are constructed for; and while reliability tends to be high, validity might suffer as the assumption that words in a dictionary have exactly one meaning is seldom met (Boumans and Trilling, 2015; Loughran and Mcdonald, 2011). Additionally, problems often arise if negations, sarcasm and irony are not accounted for.

A category of approaches located in the middle between those extremes is *supervised machine learning*. While there are major differences in how supervised learning algorithms work, the basic idea behind them is the same: first, human coders select categories and code a set of texts. Then this sample is fed to a computer in order for an algorithm to "learn" how to code texts to subsequently code the remaining texts into the determined scheme. Machine learning thus has an inductive as well as a deductive phase: the deductive part is when the researcher provides training data, which was classified using knowledge about prior research as well as employing human understanding of the meaning of the language in a text. However, the researcher does not provide explicit rules for how to look for categories — as would be done by a dictionary. Instead, the inductive phase of the process is that the computer defines a set of rules or, more generally, the decision criterion, automatically based on the training data. Simply put, the computer uses the vocabulary of documents in a known class and compares it with the words that make up a previously unclassified text. If they are sufficiently similar, this class is assigned (Jurafsky and Martin, 2020, pp. 55–75). Since this process is more effective and often outperforms dictionary methods (Boumans and Trilling, 2015), *supervised machine learning* is deemed the optimal approach for this thesis. The validity of supervised machine learning is also relatively easy to establish as, unlike fully automated methods, the expectation is that they produce essentially the same results as a human (Chang et al., 2009; Maier et al., 2018). How exactly validation is performed is discussed in Section 4.2.4.

4.2.3 Semi-Automated Content Analysis of Media Frames — The New Approach

The review of currently available approaches to *identify* and *code* media frames on a larger scope retrieved two valuable procedures, which — to my knowledge — have not been combined before: dimension reduction of frame elements to *identify* frames, as introduced by Matthes and Kohring (2008), and supervised machine learning to *code* frames. To recap, the main principle introduced by Matthes and Kohring (2008) is that frames are basically recurring patterns of different frame elements used together in a text. This means they can be seen as latent dimensions, making dimension reduction an appropriate approach to uncover these patterns. SML methods are included since they enable analysis of a larger dataset by learning coding patterns and assessing previously uncoded text quickly and with minimal cost. Since they are designed to emulate human coding, validity can be checked by comparing their performance directly against human understanding of the framing in articles.

In Figure 4.1 I present my new method for the analysis of media frames that combines these two procedures. The development and application of the new method constitutes one of the contributions of the thesis and importantly enables me to conduct the analyses that will answer the research questions about media framing of protest. I thus discuss the steps of the method both in general terms and in light of the work in this thesis. On the left side of Figure 4.1 are the individual steps, while the boxes in the centre contain the object of each step. Arrows display the connection between steps. Step 1–5 in the figure show the approach to *frame identification*, which is the inductive phase of the procedure, closely following Matthes and Kohring (2008). In Step 6–8, the results from the manual coding effort are used to train a model which codes the remaining data. This deductive part of the procedure is from here on referred to as *frame coding*.



**The actual number of frames varies between projects*

Figure 4.1: Frame Analysis Method

Specifically, Step 1 in Figure 4.1 consists of finding potential codes in the literature. As mentioned, I use the literature on the *protest paradigm*, discussed in Section 2.2.2, and the material itself. The method is not strictly sequential, which means that new codes from the material can be added to the codebook at any time. In Step 2, the codebook is constructed, using the frame elements by Entman (1993) and the extension by Matthes and Kohring (2008), who divide frame elements into codes for content analysis (see the codebook in Section 6.1). Step 3 consists of coding the material. In this thesis, I used a random sample of 500 articles drawn from the dataset of newspaper coverage of protest events in the UK which is presented in Chapter 5. Each coded unit has several binary categories, one for each code, on which they can score either 1 if a certain code is present or 0 if it is not. This results in a matrix with documents as rows and codes as columns ($documents \times codes$). The results of manual coding should be validated to establish the reliability of the results between coders. How this is done is described in the next section. In this thesis, I compared the coding of three coders. In Step 4, the dimensions of this matrix are reduced using a suitable method. This can be done with various methods but in this thesis, I use cluster analysis, as suggested by Matthes and Kohring (2008), and factor analysis. Through this, a previously undetermined number of frames emerges in Step 5. Both of the mentioned dimension reduction techniques also offer information on class membership for each individual newspaper article.

Since a set of documents with class membership is now available, I can use the same coded data to train several supervised learning algorithms in Step 6. For this, I first divide the manually coded articles into *training* and *test* set. The models are evaluated by checking the predicted classes for the test set against the actual classes as described in the next section. If the reliability metrics show a poor fit, the procedure can jump back to one of the previous steps by changing the algorithm, processing the data differently, or going further back to increase the size of the manually coded sample. Up to and including Step 6, I work with a relatively small random sample from the dataset of news media articles on protest. Regarding the size of the manually coded sample, tests have shown that after increasing the training set to a size of about 500 documents, the models only get incrementally better by adding more hand-coded documents to the training set (Hopkins and King, 2010). However, they *do* get better and in cases in which model fit is poor, the training data can be increased. Once the models are validated, the remaining data is coded in Step 7, which results in a complete set containing information of frame presence in each article in the dataset.

Note that, theoretically, instead of using supervised machine learning to replicate frames on the entire set in Step 7, it could also be used as Step 4 to determine the presence of each code in each article. In this case, dimension reduction could be performed on data from the entire dataset instead of the sample. As mentioned above, it is easier for humans to code pieces of a frame instead of coding frames holistically. However, Burscher et al. (2014) establish that the same is not true for computers. The problem often identified for manual coding is that humans lack the mental capacity to evaluate the entire text at once and will therefore focus on smaller portions of a document, which leads to subjective decisions in coding tasks as they are too demanding. Computers, however, do not have this limitation as they can evaluate a plethora of information at once. In fact, the computer will do so no matter if the task is to code a holistic frame or a frame element since the basis for decisions is the entire vocabulary of the text each time. In this case, training models for each frame element hurts validity as errors in the coding of each element add up. So while humans are better at coding frame elements, computers are thought to be better at replicating coding of entire frames.

In theory, this approach allows a few different units of analysis: the whole article could be employed which maximises the chance that each unit will have information about each of the frame elements. However, as reporters will often try to at least make their stories appear balanced to some degree, different parts of a story might stress different views, which can make coding decisions harder. The sentence level, however, is often too short to identify a whole argument which will usually stretch over at least a few sentences. It was therefore deemed most appropriate to use the paragraph as the unit of analysis during manual coding since paragraphs in journalistic texts usually comprise just one thought or argument but also have sufficient information for valid coding. As it later turned out, the computer works better on the article level though. The reason is likely the same as above: computers have the ability to process larger amounts of information at once. After manual coding, paragraphs were, therefore, collapsed and values for each code represented the article mean from all paragraphs.

To sum up, the described method of frame *identification* and *coding* that is used in this thesis solves several issues. By employing Matthes and Kohring (2008) approach of frame *identification*, it is both firmly based on theoretical underpinnings of the framing concept and makes decisions and analysis steps more transparent than other approaches. Frames are not presented as given or appear from a pilot of close reading by different researchers — which ultimately remains a black box for all but the involved. Instead, they emerge from the coded data. The individual coding decisions are markedly easier than coding frames holistically as coders only need to choose between small clearly defined codes based on the idea of frame elements. Once frames are identified and the presence of each frame is determined in each of the coded articles through dimension reduction, ACA, and specifically supervised machine learning, is used to replicate insights into the manually coded sample on the entire dataset. This allows for the broad scope deemed necessary to answer the posed research questions. However, coding frames in the dataset of news media coverage of protest ultimately serves to provide the dependent variable for the analysis in Chapter 7 as well. The independent variables for that analysis are explained in Section 4.3.

4.2.4 A Note on Sound Measurement

In content analysis, as in other research, the soundness of measurement is crucial. In the method introduced in the last section, there are two potential pitfalls for the reliability and validity of results: during the manual coding performed in Step 3 and during the computational coding performed in Step 7. For both steps, establishing the soundness of measurement works essentially in the same way: a part of the measurement is done independently by two or more coders who assess the same material to test whether they reach the same result.

For manual coding, this is done by different researchers who use the same instructions while coding the same material to create *reliability data*. Afterwards, their agreement is assessed to determine if the process can be deemed reliable. In other words, a data-making procedure is deemed reliable if it can be reproduced (Krippendorff, 2004, pp. 211-221). To eliminate the chance that coders arrive at the same conclusion by chance or by a flaw in the research design, Cohen's κ (kappa) (Cohen, 1960) and Krippendorff's α (alpha) (Krippendorff, 2004) are usually calculated to establish *reliability* for manual coding.

Cohen's κ is defined as:

$$\kappa = \frac{A_o - p_c}{1 - p_c} \tag{4.1}$$

In this equation, A_o is the proportion of units with matching categories among the units coded by both coders, usually called agreement. p_c is the agreement that can be expected by chance. Ideally, p_c would be based on the real proportion of the categories in the population. Since knowledge about the real proportion is usually not available in content analysis, however, p_c is calculated from the number of times all coders assign each category. In this thesis, I only calculate reliability between two coders, A and B. For this case, agreement expected by chance is defined as:

$$p_c = \sum_k \frac{n_A k}{N} \cdot \frac{n_B k}{N} \tag{4.2}$$

k here means the number of categories. Since I only use binary codes in this thesis, the formula can be simplified as follows:

$$p_c = \frac{n_A0}{N} \cdot \frac{n_B0}{N} + \frac{n_A1}{N} \cdot \frac{n_B1}{N} \tag{4.3}$$

n_A0 is the number of times coder A uses category 0, n_B1 is the number of times B uses 1 and so on (Cohen, 1960).

Krippendorff developed the α-agreement with the same goal as Cohen (1960), namely to correct the impact of chance in the agreement between coders. Krippendorff's α is more flexible and can account for different sample sizes, missing data and more than two coders (Krippendorff, 2004, pp. 244-250). While these features are not crucial for this thesis, I report both metrics to triangulate the different approaches.

Krippendorff's α is defined as:

$$\alpha = 1 - \frac{D_o}{D_e} \tag{4.4}$$

In this equation, D_o means the *disagreement* between coders, while D_e is the *disagreement* which can be expected by chance. Calculation of D_o and D_e can quickly grow complex as the number of coders and categories grows (see Krippendorff, 2004, pp. 230-232). Therefore, I only provide the definition for the case relevant in this thesis here — two coders and two categories:

$$\alpha = 1 - \frac{o_01}{n_0 \cdot (\frac{n_1}{n-1})} \tag{4.5}$$

Where o_01 is the number of times coder A and B disagreed, n_0 is the number of times 0 was assigned by any coder, n_1 is the number of times 1 was assigned by any coder, and n is the number of times any code was assigned by any of the coders (Krippendorff, 2004, pp. 221-227).

For machine coding, the process is almost identical, yet with an important conceptual difference: Cohen's κ and Krippendorff's α compare human coders who are all potentially correct in their judgement of an otherwise impossible to know class of a document; the machine learning literature assumes the class assigned by human coders represents the *true* one and only the performance of the machine is evaluated. The reliability of ACA methods tends to be perfect as computers will always return the same result given the task to perform the same analysis on the same data. However, as Grimmer and Stewart (2013) put it, “all quantitative models of language are wrong — but some are useful”. The *validity* of results can not be assumed, as computers do not understand human language as such but deal with simplified models of language based on word co-occurrence and counting.

As said, establishing the soundness of measurement is similar, despite this conceptual difference: after coding a portion of the data manually, the coded material is assessed again by the computer — just as if it was just another coder. In practice, the manually coded sample is randomly divided into two sets. A larger sample called *training set*, which is used to *train* the supervised learning algorithms on how to code texts, and a smaller *test set*, which the computer treats as new data and predicts class membership. The predicted class of the *test set* are then compared to the human coding. To establish agreement in machine learning, four categories are important:

- *true positive*, meaning that the computer correctly predicts that a case is positive;
- *true negative*, meaning that the computer correctly predicts that a case is negative;
- *false positive*, meaning that the computer incorrectly predicts that a case is positive;
- *false negative*, meaning that the computer incorrectly predicts that a case is negative.

Table 4.1 shows these definitions in a confusion matrix, which is how (dis)agreement is usually presented in the machine learning literature.

Table 4.1: Confusion Matrix

		manual	
		0	**1**
computer	**0**	true negative	false negative
	1	false positive	true positive

Similar to Cohen's κ (kappa) and Krippendorff's α (alpha), the machine learning literature uses measures that make sure a high agreement between coders is not simply due to chance. These are *precision*, *recall* and their harmonic mean, referred to as *F1-score* (also see Grimmer and Stewart, 2013; Manning et al., 2008, pp. 142-144). Precision is defined as:

$$\text{precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}} \tag{4.6}$$

This means that precision represents the proportion of machine-predicted values that are correct. However, precision should be accompanied by recall, as a model which always predicts positives has the same precision as the proportion of positive values in the population. Recall is defined as:

$$\text{recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}} \tag{4.7}$$

Recall is thus the proportion of the human coded values that are predicted correctly. Often, models are optimised for either precision or recall. If a model is supposed to detect cancer, for example, false positives are not nearly as severe as false negatives. In this case, models should have a recall as high as possible. On the other hand, if one is looking for a picture of an ugly dog, the model should be optimised for high precision, as seeing a pretty dog instead, which is the vast majority of the dog population, should not be too troublesome.

The *F1-score* formalises this trade-off and puts equal weight on both scores:

$$F_1 = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{4.8}$$

Formally, the F_1 score is just a special form of F_β scores where recall is considered β times as important as precision. However, since precision and recall are usually equally important, the F_1 score is much more commonly used than other F_β metrics (Manning et al., 2008, pp. 142-144).

4.3 Independent Variables

To answer the hypotheses formulated in Section 3.4, eight independent variables are used: *Protest Goal*, *Protester Violence*, *State Repression*, *Newspaper Ideology*, *Newspaper Type*, *Ideological Divide*, *Days Since Start*, and *Year of Protest*. As mentioned above, the thesis employs two datasets. One consists of newspaper data retrieved from the archive of *LexisNexis* and contains information about two of the independent variables, namely *Newspaper Ideology* and *Newspaper Type*. For the remaining variables, a second data source was needed. This section first discusses the choice for a protest event database and discusses several potential alternatives which were discarded. The second part of the section then briefly explains how the media data was matched with the protest event data. More information on how the individual independent variables were processed for analysis are discussed in Section 7.1.

4.3.1 Data

The analysis below employs three kinds of independent variables: event-level, outlet-level and time-bound variables. Some of these are already present in the dataset of news coverage described in Chapter 5. However, several question, especially on the event-level, are not as easy to answer: When did a protest occur (i.e., when did it start and finish)? What did the protesters want? Did they engage in violence? And how did the state react to a protest? This information is not present in the downloaded newspaper data. At least not directly. It would be theoretically possible to read the articles and extract the information from the reports manually, as has been done by several similar studies (e.g., Kilgo and Harlow, 2019; Lee, 2014; Shahin et al., 2016). However, given that this research emphasises scope and covers a long time period instead of focusing on a specific protest or movement, this strategy was not considered feasible.

Instead, several available protest event databases were considered. Protest event data are predominantly collected and used by researchers who study social movements, basically since the 1960s. The goal of these databases is to code, in a standardised way, at least the basic features of events, specifically, the "who, what, when, where, and why of the event" (Earl et al., 2004, p. 72). Traditionally this data was collected primarily in sociology using mostly manual coding of media accounts of protest events to study social movement dynamics (Tilly, 2002). Only more recently researchers and companies started to use automated methods to measure and forecast political conflict (Fisher et al., 2019).

As Earl et al. (2004) point out though, using media accounts to collect information about protest leads research to repeat the selection bias and description bias described in Section 2.1. Research using media-independent information about events has shown, however, that publishing outright false information about an event is relatively rare (McCarthy et al., 1999). Reports rather omit information or misrepresent what happened during an event through framing (Earl et al., 2004) — like emphasising, for example, violence even when violence was only a very small aspect of what happened in the course of an event, as discussed in Section 3.3. In addition, the goals and grievances of a protest are more likely to be omitted from reports which focus on arrests, violence and characteristics deemed more interesting (Smith et al., 2001). Apart from goals, reports about the variables in this thesis are generally considered accurate. Bias can, however, be expected when several sources report conflicting information about an aspect of an event: Especially the crowd size, which is not considered in this thesis, might be a problem here as authorities often have incentives to downplay the numbers while protesters would prefer higher numbers to be reported (Oliver and Meyer, 1999). To minimize this problem, researchers building those databases are advised to triangulate multiple sources and usually do so (Earl et al., 2004).

Protest event databases are, therefore, thought to reproduce media bias if they rely on media accounts of what happened during an event. However, media-independent information about events is often scarce and might be equally biased[11]. Data coded from news reports thus still seems to be the best option.

Besides these general considerations, this research places some practical requirements on a database: it must cover protest events that occurred in the UK without systematic omission in the chosen time span (1992-2017) and provide data for the independent variables listed above. Drawing on Fisher et al. (2019), it becomes clear that most of the available protest event data is, therefore, not suitable for this project: The *Dynamics of Collective Action* (McAdam et al.) project only gathered data for US protest and only until 1995; Francisco (n.d.) gathered data for protests in 28 European countries, including the UK, yet only from 1980 until 1995; Salehyan et al. (2012) collected data on social protests in Africa and Latin America; the *Mass Mobilization in Autocracies* data project (Weidmann and Rød, 2019) logically excludes the UK as it is not an autocracy.

Recent approaches use computerized natural-language methods to parse the text of media and social media data to produce event data. Notably, the Count Love and the Crowd Counting Consortium (CCC) projects have started to collect data about protests based on reports in news articles, social media posts, advocacy groups announcements, and attendee submissions in near real-time (Fisher et al., 2019; Leung and Perkins; Pressman and Chenoweth). However, both projects only started their work in 2016. The most comprehensive database to date is the Global Database of Events, Language, and Tone (GDELT) project (Leetaru, 2020) which monitors different media types in over 100 languages for certain event types, including protest. It also

[11] McCarthy et al. (1999), for example, use police records, which arguably suffer from other biases.

contains information going back to 1979. However, while it provides some metadata on the protest events like geolocation and number of protesters, it misses many variables identified above, especially protest goal and tactics.

More recently, the Observatory for Political Conflict and Democracy (*PolDem*) project started to provide three datasets on protest events: “PolDem-Protest Dataset 30 European Countries” (Kriesi et al., 2020b), “PolDem-Protest 6 European Countries” (Kriesi et al., 2020a) and “PolDem-Protest Dataset on EU issues” (Grande et al., 2020). However, all three are also media-based: Kriesi et al. (2020b) use ten English-language newswire agencies and *LexisNexis*; the codebook for Kriesi et al. (2020a) does not list their collection strategy. Grande et al. (2020) use one quality newspaper per country, which is *The Times* for the UK; the dataset by Kriesi et al. (2020b) is probably the most thorough one since it uses an automatic approach to sift through media data from outlets in different countries to combine the insights. Together the three datasets cover the entire time frame, yet since the quality and focus of the data is different, the number of events they find per year differs considerably, as shown in Figure 4.2. In terms of variables, all three datasets coded some of the information required for the independent variables mentioned above. However, tactics/violence was only coded by Kriesi et al. (2020b) and the state response was only coded by Grande et al. (2020).

After this thorough scoping, the conclusion is that the only database available at the moment that covers protests in the UK for the entire time frame and includes all key variables is the *Mass Mobilization Project* (Clark and Regan, 2019). However, like other projects, the *Mass Mobilization Project* still relies on media accounts of protest events. Specifically, Clark and Regan (2019) use the keywords “Protest”, “Demonstration”, “Riot”, and “Mass Mobilization” on the *LexisNexis* newspaper archive, selecting “All News” as the source. For the UK, the *Mass Mobilization Project* lists 471 events between 1992 and the end of 2017, as shown in Figure 4.3.

Using the *Mass Mobilization Project* data, or in fact, any of the mentioned datasets, thus creates a substantial limitation for this research: since the dependant variable

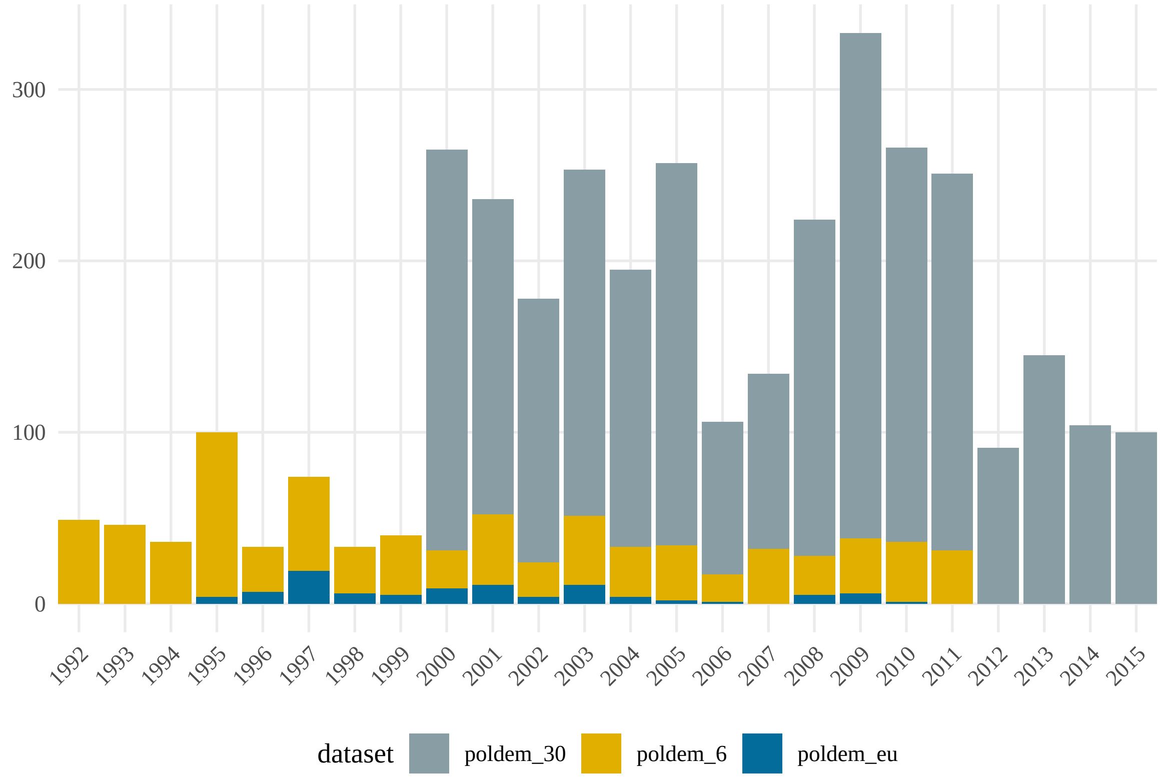


Figure 4.2: Number of Events in the PolDem Datasets over Time

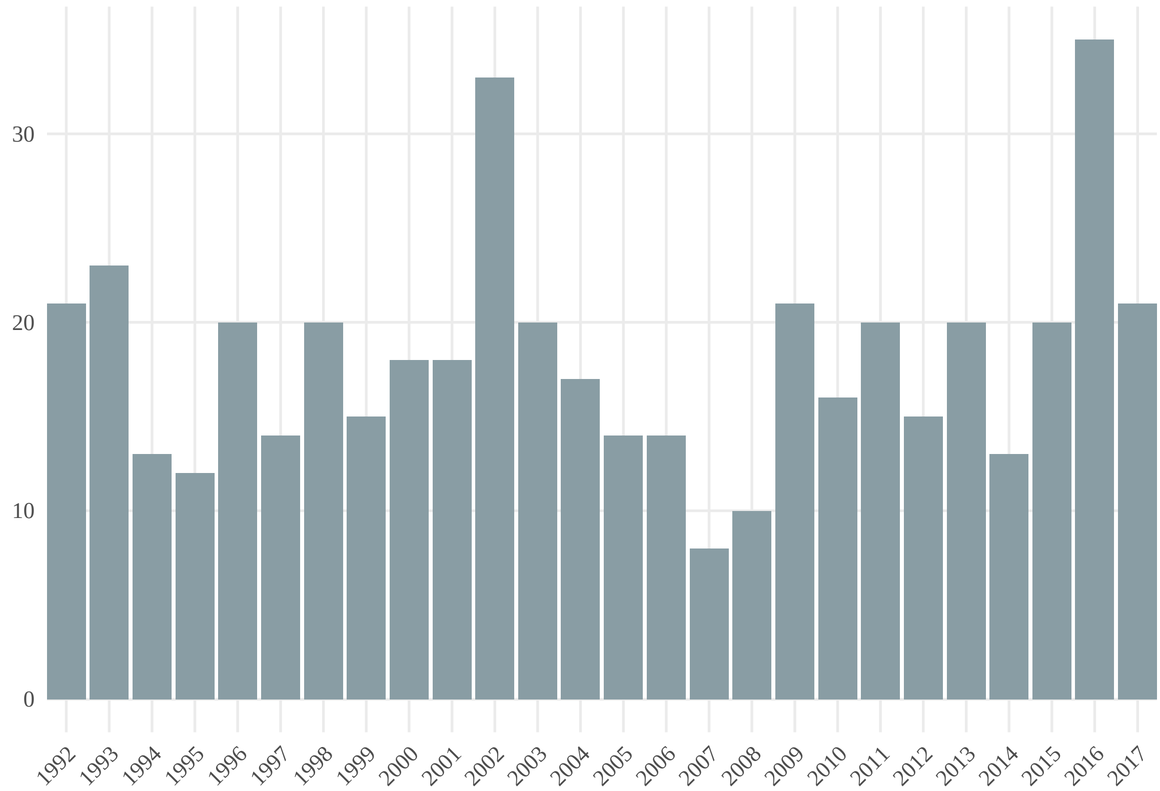


Figure 4.3: Number of Protest Events in the UK per Year per Clark and Regan (2019)

concerns how a protest is covered in newspaper articles, and the independent variables were collected from newspaper data as well, endogeneity is potentially an issue for this thesis. When it is scrutinised in Chapter 7 if violence during an event leads to more delegitimising coverage, for example, both the information about whether violence occurred and if the violence was disproportionally highlighted by the media are based on the same data source. It was possible, however, to use one mitigation strategy: the *Mass Mobilization Project* used several different newspapers to code data in the UK. Among these, I only used *The Times* to construct the dependent variable — media framing of protests. As a robustness test, all analysis steps involving the *Mass Mobilization Project* dataset were performed again without data from *The Times*. The models produced with this altered dataset showed no substantial differences, confirming the robustness of the results. Clark and Regan (2019) also note which specific articles they used for coding. These were removed altogether from the newspaper dataset. Nevertheless, the data I use for the variables *Protest Goal*, *Protester Violence* and *State Repression* come with this important caveat.

Furthermore, the *Mass Mobilization Project* (MMP) exacerbates another problem: Clark and Regan (2019) only collect data for protests with at least 50 observed participants. This makes the selection bias already criticised by social movement scholars (Earl et al., 2004) more severe. Not only does it mean that protests which are not covered are absent in the database, the data is also biased towards higher profile protests which attract more attendees. Nevertheless, since this dataset is the only viable alternative besides creating a dedicated new dataset — which would be excessively difficult and time consuming given the obstacles described above — the additional limitations introduced to this thesis seem a necessity. Specifically, the great advantage of the MMP data is that I can analyse the media framing of a wide range of different protest events over the entire 26-year period, which would not otherwise be possible. As shown in Figure 4.3, there is also no noteworthy trend in the number of protests over time. So an existing selection bias is at least stable over time, meaning that conclusions about changes in framing over time should remain unaffected.

4.3.2 Matching Strategy

After getting access to the MMP data, the challenge was to match the protest events with my news coverage dataset. Which articles belong to which event? To merge the two datasets, I employed a two-staged strategy: first, I selected all articles in the newspaper dataset that were published no earlier than seven days before an event started and no later than 30 days after it ended; then I used the variables "Protest Locations", "Protest Group Identity" and "Protester Demands" to extract suitable keywords (see Appendix D). I then used keywords to filter the articles selected in the first stage: articles were discarded if they did not feature at least one of the keywords. This selection was manually validated and overlaps between events were separated by hand.

After removing the articles from *The Times* which were used for coding event data, some events could no longer be linked to any articles in the database. Additionally, I removed several protest events which did not meet the definition of protest as described in Section 4.2.1.[12] For many articles, it was not possible to match them to any event in the MMP data. After the matching procedure, 5,639 of the total 27,496 newspaper articles remained. Likewise, from the 471 events in the *Mass Mobilization Project* which took place in the UK, 341 remained after the data had been processed and matched.

Chapter Conclusion

This chapter has outlined the research design, data and methods employed in this thesis. The first section discussed why the specific case was chosen. To capture the expected changes in reporting patterns, I chose a 26 year period starting in 1992. This

[12] The excluded cases were acts of sectarian violence and rioting in Northern Ireland, which did not provide a specific goal or grievance, or industrial action. These were excluded.

year marks the start of influential technological developments linked to the emergence of the internet as a mass phenomenon. To circumvent the limited scope of previous studies, which often studied rather specific protest events, I decided to include all coverage of domestic protest in mainstream UK newspaper in the study. Finally, the choice for the UK was primarily influenced by considerations of the two independent outlet level variables of the research: newspaper type and newspaper ideology. The media economy in the United Kingdom is traditionally marked by partisan news outlets and a rather uniquely sharp segmentation between broadsheet and tabloid newspapers. This means that the influence of both variables can be expected to be more pronounced in the UK.

The second and third part discussed the dependent and independent variables of this thesis, which also reflect the two-step procedure of analysis. Specifically, the operationalisation and measurement of the dependent variable — media frames in protest reporting — is one of the main goals of this thesis. Section 4.2 defined what is understood as protest in this thesis before surveying available approaches for manual and automated content analysis of media frames. I then outlined the new approach developed for this thesis which will help to perform a large scale framing analysis on the coverage of protest in the UK. Once this information is available, three of the hypotheses summarised in Table 3.1 can be answered.

The last part discussed the independent variables used in the second analysis step. Scrutinising available protest event databases, I found that none of them was a perfect match for this thesis. Especially the reliance on media data for the coding of event information means that the research will inevitably face endogeneity problems. However, since compiling a new dataset was deemed unfeasible, the *Mass Mobilization Project* database (Clark and Regan, 2019) was found to be the best choice. The final sub-section of this chapter discussed the matching strategy used to combine the news media dataset in which media frames are coded with the protest event database by Clark and Regan (2019).

Chapter 5

Building a Dataset of News on Protest Events

A key goal of this thesis is to gain a more systematic understanding of how the media reports about protest. To do so, I use an inclusive data collection strategy that aims to include all reports about any domestic protest in newspapers in the United Kingdom since 1992 — which fall under the definition of protest in Section 4.2.1. This chapter describes how the news media data for this thesis was gathered and how the resulting dataset was cleaned to include only relevant articles.

The chapter is divided into four sections. The first discusses the corpus of interest, or more specifically, why the newspaper outlets *The Mirror*, *The Sun*, *Daily Mail*, *The Guardian*, *The Independent*, *Financial Times*, *Daily Telegraph*, and *The Times* were included. In the second section, I discuss the employed newspaper archive, *LexisNexis*, and highlight some problems specific to this service and digital news archives in general. One potential problem is the omission of relevant news items during data retrieval. The third section discusses a pilot study I conducted to avoid this problem by using natural language processing (NLP) techniques to find a comprehensive list of keywords for data retrieval. Using these keywords, I downloaded a large corpus of newspaper articles.

In the remaining section, I discuss how this corpus was subsequently cleaned from irrelevant data to obtain a suitable base for further analysis. In hindsight, not all steps appear to have been necessary, although all had been successful in that they made the dataset cleaner. Nevertheless, describing them is hopefully useful for other researchers facing a similar task: collecting a database of news media stories about a concept which is used in several contexts — most of which are *not* relevant to the topic of research.

5.1 Newspaper Selection

As mentioned above, one of the reasons why the UK was selected as a case was the specific nature of its newspaper market. While Tuchman (1972) has argued that objectivity is a strategic ritual for US reporters, British press outlets traditionally have their distinct partisan identity (Hallin and Mancini, 2004; Kuhn, 2007). More specifically, while the broadcasting sector traditionally stands under close public and political scrutiny when it comes to the quality and "objectivity", or at least balance, of its reporting, the press is largely self or essentially unregulated (Hallin and Mancini, 2004; Kuhn, 2007). Professionalism in British newspaper journalism is thus less tied to a notion of balance or objectivity and is rather characterised by mastering routines and standards of the profession which include pushing aside personal beliefs in favour of predefined editorial lines and the identity of the organisation (Hallin and Mancini, 2004).

This means that by selecting tabloid and broadsheets as well as left- and right-leaning outlets in the UK, it should be possible to maximise the differences in reporting caused by ideology (as theorised in ***H3e-h***) and newspaper type (***H3i-j***). The choice for newspaper outlets is shown in Table 5.1 and reflects the aim to cover the variance in these two key variables. Additionally, outlets with a higher circulation were preferred in order to analyse the content that is most likely to have been read. As discussed above, time likely also explains a fair share of variation which meant that the availability of outlets in the employed archive, *LexisNexis*, was essential.

Table 5.1: Selection of Newspaper Outlets for Data Collection

Type	Ideology	Outlet	Available from[a]	Circulation (2005 – 2017)
Tabloid	Left	The Mirror	1995	1,720,000 – 692,295
Tabloid	Right	The Sun	2000	3,273,000 – 1,602,320
Tabloid	Right	Daily Mail	1992	2,426,000 – 1,442,924
Broadsheet	Left	The Guardian	1992	366,645 – 153,431
Broadsheet	Left	The Independent	1992[b]	263,595 – NA
Broadsheet	Right	Financial Times	1992	419,386 – 190,046
Broadsheet	Right	Daily Telegraph	1992	460,585 – 907,329
Broadsheet	Right	The Times	1992	679,190 – 440,736

[a] Source for circulation numbers: http://www.magforum.com/papers/nationals.htm

[b] No figures available for The Independent in 2017 as the outlet moved to online-only publication in 2016.

5.2 A Note on Digital News Archives and *LexisNexis*

Digital news archives are an indispensable tool for most media studies as the alternative would be the time and resource-intense gathering and digitalisation of old issues. For this thesis, I use the newspaper database of *LexisNexis* to retrieve articles on protest using a keyword search. Among social scientists, *LexisNexis* is — or at least has been for a while — the most often used data source for newspaper analysis (Deacon, 2007; Weaver and Bimber, 2008). This is also illustrated by the fact that many of the studies reviewed in Chapter 2 employed *LexisNexis* (e.g., Boyle et al., 2012; Dardis, 2006a,b; Di Cicco, 2010; Kyriakidou and Olivas Osuna, 2017). *LexisNexis* makes it possible to generate an advanced search string by combining multiple key terms with Boolean search operators[13] to extend or restrict the range of a query. This makes it possible to select archived articles by content, in addition to narrowing selection by dates and outlets.

However, some people have also noted limitations of *LexisNexis* as a data source in terms of reliability and validity of research, which need to be considered. Firstly, while the database is electronic and article retrieval should have a high reliability, compar-

[13] The three basic boolean operators are: *AND*, *OR*, and *NOT*. They can be used to combine search commands in databases to broaden or narrow down results.

isons between *LexisNexis* against other databases and hard copy issues of newspapers have shown sporadic inconsistencies (Deacon, 2007; Weaver and Bimber, 2008). Most likely, this is due to human errors during the data entry phase, but since the data collection and retrieval process is basically a black box for the user, it is hard to systematically uncover errors.

A second issue, which underlines this impression, is the presence of duplicate data. Deacon (2007) mentions that multiple entered articles are almost always present in data retrieved from *LexisNexis*. Some of these duplicated articles are apparently added as different editions of the same outlet (e.g., regional editions). Yet, since information about the edition is not consistently provided, it is often hard to tell why articles appear twice or even several times in the retrieved data. It can also often be observed that articles appear to be duplicated at first but show minimal differences on closer inspection. The phenomenon is made even more problematic as it is more common for some outlets than for others, which might distort coverage patterns. However, this issue is not as problematic as before, since automated tools exist now to remove duplicated entries from the dataset (Gruber, 2021).

There are other concerns regarding the validity of research in digital news archives: articles are archived without any visual dimension — such as graphics and placing within the newspaper (Weaver and Bimber, 2008). This means that what the researcher gets to analyse can be somewhat different from the experience a reader had at the time. However, since the primary focus of this research is patterns in content rather than the respective content's effect on readers, the impact of this limitation should be minimal.

In contrast, a different issue in terms of validity is highly relevant to this research: using keywords to retrieve a subset of the available content — which is one of the main advantages of a digital archive — means that the choice of the correct list of keywords becomes crucial. Choosing keywords that are unfit to capture the concept or theme that is supposed to be analysed might omit relevant articles and retrieve irrelevant ones instead. Since this is the basis for further analysis, mistakes during this step can skew all subsequent results. A useful concept in this regard is the distinction between *false positives* and *false negatives*, already presented in Section 4.2.4 (also see Soothill and Grover, 1997).[14]

In this case, *false positives* are articles which were retrieved because they mention at least one of the keywords, yet do not capture the concept of interest. This is often due to the inevitable ambiguity of keywords. For this research, the keyword "protest", for example, is one obvious choice and has been used by several of the reviewed studies (e.g., Boyle et al., 2012; Dardis, 2006a,b; Di Cicco, 2010; Lee, 2014). Yet, there are a number of sentences featuring this keyword that have nothing to do with the public gathering of protesters. To give a few examples which came up while constructing the database: "Prince William, we are told, noticed the camera and telephoned his father to protest"; "Her departure from the newspaper provoked many protests from readers."; "A vote for Paddy is the perfect protest vote". In previous studies, this might not have mattered so much as manual coding provides an opportunity to filter out false positives at a later stage. However, with a large number of documents in this dataset, a manual review of the relevance of all newspaper articles is unfeasible. As described in Section 5.4, I solve this problem by employing machine learning techniques to filter out irrelevant articles, as suggested by, for example, King et al. (2017).

A second problem are *false negatives* — i.e., potentially relevant articles which are not captured by a keyword search (Soothill and Grover, 1997). Missing articles is a more severe problem here, as one usually does not even know what, if anything, is missing. Using keyword searches can make this problem more severe if the chosen keywords do not capture the concept of interest appropriately (Deacon, 2007). This issue is exacerbated by the well-established fact that humans perform exceedingly poorly at the task of identifying appropriate keywords for search queries due to limitations of the human brain (Bäuml, 2008; King et al., 2017). To solve this problem, I follow

[14] The concept is mirrored by Krippendorff's (2004) distinction between errors of omission (the failure to retrieve relevant texts) and errors of commission (the retrieval of irrelevant texts) (p. 276).

suggestions to pilot a search query with as many potentially relevant keywords as possible before reducing the sample to contain a more parsimonious set of keywords which can still retrieve the entire population of interest (King et al., 2017; Soothill and Grover, 1997). The details are explained in the next section, before Section 5.4 describes the removal of false positives.

5.3 Pilot: Finding the Right Keywords

The main goal of the pilot was to identify keywords to use for a search on the *LexisNexis* newspaper archive. In short, the strategy of the pilot was to create an initial list of keywords related to protest which was as broad as possible, to prevent false negatives. A sample was then downloaded from *LexisNexis* and the keywords were evaluated on their ability to retrieve relevant results, to minimise false positives.

As the terms to describe protest might change over time and between different outlets, a sample was drawn from all outlets and within the whole period (1992-2017). Queries were then formed to retrieve articles using any of the keywords in a constructed week each year.[15] To identify potential keywords, four sources were consulted with the intention to find *any* word which had even the slightest relevance to the topic, since omission of relevant terms means that respective articles using only this term would be lost forever: Tilly's book *Contentious performances* (2012), which maps a number of different means groups use to make public statements; Della Porta's chapter on *Repertoires of Contention* (2013); the Wikipedia page about protest (Wikipedia, 2020); and a search for synonyms of "protest" and "demonstration" on the webpage Thesaurus.com (Thesaurus.com, 2018). Browsing all available sources for potential keywords is a good strategy as humans are rather bad at coming up with more than a couple of terms on their own but are exceedingly competent at deciding which terms are relevant (King et al., 2017). This resulted in a list of 35 keywords, which are shown on the y-axis of Figure 5.1. To catch all variations of keywords, an exclamation mark was added to

[15] One randomly selected Monday, Tuesday, Wednesday and so on from each year.

keywords where this made sense.[16] Searching for these 35 keywords on the randomly selected days produced 103,570 results for 182 days (i.e., seven per year). However, these articles were, as it turns out, mostly irrelevant to the study of protest coverage.

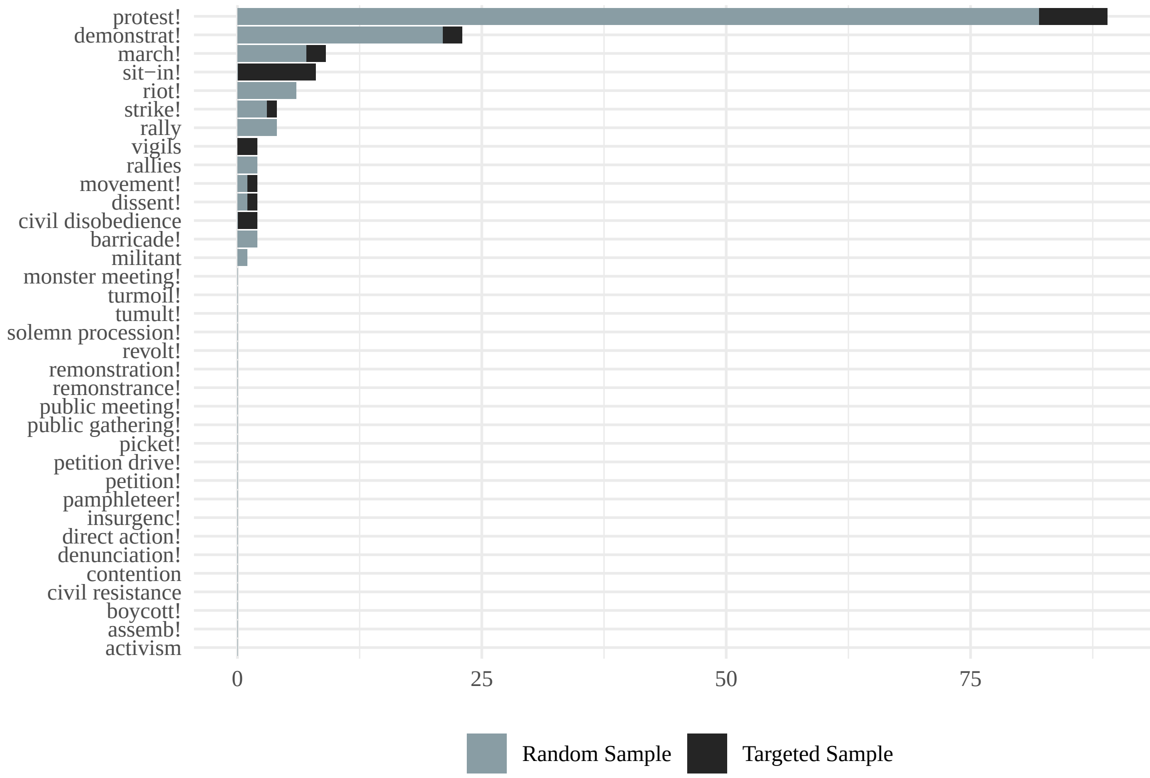


Figure 5.1: Number of Times Each Keyword Was Relevant in the Pilot Sentences

To establish which keywords retrieved relevant results, two samples were coded. First, all articles were processed and split into sentences using the R packages *LexisNexisTools* (Gruber, 2021) and *quanteda* (Benoit et al., 2018). Sentences containing none of the keywords were discarded. A first sample of 1,000 sentences was chosen at random and coded manually for relevance: did they speak about protest as defined in Section 4.2.1 or not? The result showed that the large majority of the sentences, 89.5%, was irrelevant. Coding showed that several of the keywords were regularly used in other contexts — which does highlight the importance of a pilot to construct a search string. "March", for example, was one of the most often found keywords, yet 172 of the 228

[16] On *LexisNexis* searching, for example, for "protest!" retrieves all suffixes of the word, like "protests" or "protesters". This made no sense for words where grammar does not allow extensions, such as with "rally".

mentions referred to the month. Another example is the keyword variation "striker" which was mentioned in 170 sentences about football. The keyword "strike!" was, therefore, the word most often used in irrelevant articles with a ratio of 270 to just three.

The first sample, however, had one practical issue: several of the keywords did occur in only very few instances or not at all in the original random sample. A more targeted sample was, therefore, generated to make sure that rarely employed keywords were assessed in at least ten sentences each, which lead to 127 additionally coded sentences.[17] Again, 90% of sentences turned out to be irrelevant.

Coding the targeted sample put three extra terms on the map, as can be seen in Figure 5.1: "sit−in!" (with 8 out of 10 coded sentences being relevant), "Vigils" (two relevant mentions) and "civil disobedience" (two relevant mentions).[18] Part of the coding effort also included looking for additional keywords. Yet, neither manual coding, ranking the most often used words in relevant sentences, nor performing a keyness analysis in *quanteda* (Benoit et al., 2018) led to any more relevant terms. I am thus confident that the pilot study minimised the number of false negatives as far as possible.

However, 90% of the coded pilot sentences had turned out to be false positives. As mentioned earlier, false positives are a less severe problem, since irrelevant articles can be weeded out later in the process. Yet, one should also try to reduce irrelevant data by a parsimonious use of keywords, as computational tools to remove false positives are almost never perfect. A second question during the pilot phase was thus: how many of the keywords are needed to retrieve *all* articles deemed relevant through coding? A small number of keywords is preferable, as additional keywords increase the number of returned false positives.

[17] The words "monster meeting!", "direct action!", "petition drive!", "solemn procession!", "civil resistance" and "remonstrance" had less than ten occurrences in the pilot set.

[18] Since many of the sentences mentioned more than one keyword, the total number of relevant mentions surpasses the number of relevant sentences. This also explains why the second round of coding added some more instances of the term "protest!".

To test the keywords, the original articles from which the relevant sentences were extracted were revisited. I then examined how many of them could have been retrieved using just the most relevant keyword: "protest!". Of the 116 articles coded relevant during the pilot, this query retrieved 109 (94%). This is already pretty good coverage yet would re-introduce the problem of false negatives to the sample. Adding the second most relevant term, "demonstrat!", pushed this number to 100%. In other words, using the two top keywords is ideal to accomplish the optimising task at hand: make sure to not omit any relevant articles, while not needlessly including irrelevant ones.

From the experience with the term "striker", another concern arose: what if there are variations of the two top keywords which are themselves not relevant? One instance had already been found during coding: sentences containing "Protestant" had not been coded relevant once. To assess this question, all words that partially matched the character strings "protest" or "demonstrat" were retrieved. The resulting 110 variations can be seen in Figure 5.2, which shows a wordcloud where words that appear larger were used more often in the pilot articles. "protest" was the most often used term with 5,811 occurrences while the most common form of "demonstrat", "demonstrated", was mentioned 2,339 times. To assess whether a word was relevant to the concept, 25 random sentences for each key term were retrieved from the pilot set and manually coded ($n = 714$)[19]. All words that were found to be relevant in at least one of the 25 random articles were then included in the final search. The final string, therefore, was as follows:

> "protest OR protests OR protesters OR demonstration OR demonstrations OR demonstrators OR demonstrating OR protesting OR protester OR protestations OR protestors OR demonstrator OR counter-protesters OR undemonstrative OR protestor OR enltrprotesters OR counter-demonstration OR counter-protest OR counter-demonstrators OR protester's OR demonstraters OR protestsreflect OR protest-related

[19] 90 of the 110 words appeared in less than 25 articles.

> OR counterprotests OR counter-demonstrations OR counterprotest OR demonstration's OR mini-demonstration OR eco-protests OR protest-marches OR eco-protest OR protest's OR eco-protestors OR eco-protesters OR protest-as-theatre OR counterdemonstrators OR counter-protests"

The "OR" operator, in this case, indicates that any articles that mentioned any one of the 37 variations of "demonstrat" and "protest" were retrieved. Some of the words might seem surprising, such as "protestsreflect" or "demonstraters" which are obvious misspellings. Yet, they did retrieve relevant articles and are thus fit to eliminate a few further false negatives.

5.4 Collecting and Cleaning the Database

The next step was then to download the relevant articles from *LexisNexis*, using the eight national outlets *The Mirror*, *The Sun*, *Daily Mail*, *The Guardian*, *The Independent*, *Financial Times*, *Daily Telegraph*, and *The Times* as sources and the period from January 1st, 1992 until December 31st, 2017 as the time frame. The search produced 598,962 results. This raw number still contained duplicated data and other superfluous entries, however. As with the pilot data, the first step was to get the raw files into a table format for further analysis. For this purpose, as well as other tasks in this chapter, I developed code for the statistical computing software R (R Core Team, 2021) and published it as the software package *LexisNexisTools* (Gruber, 2021).

After the data was parsed into a table format, it was possible to reduce the raw number of articles carefully and systematically in a number of steps. The heights of the bars in Figure 5.3 indicate the raw number of downloaded articles. As can be seen, the number of downloaded articles grows over time. However, only a fraction of the downloaded articles — the grey areas — remained in the final dataset, while the coloured segments represent items removed during the different cleaning steps. After cleaning, the effect of growing numbers of articles over time completely disappears. This suggests that

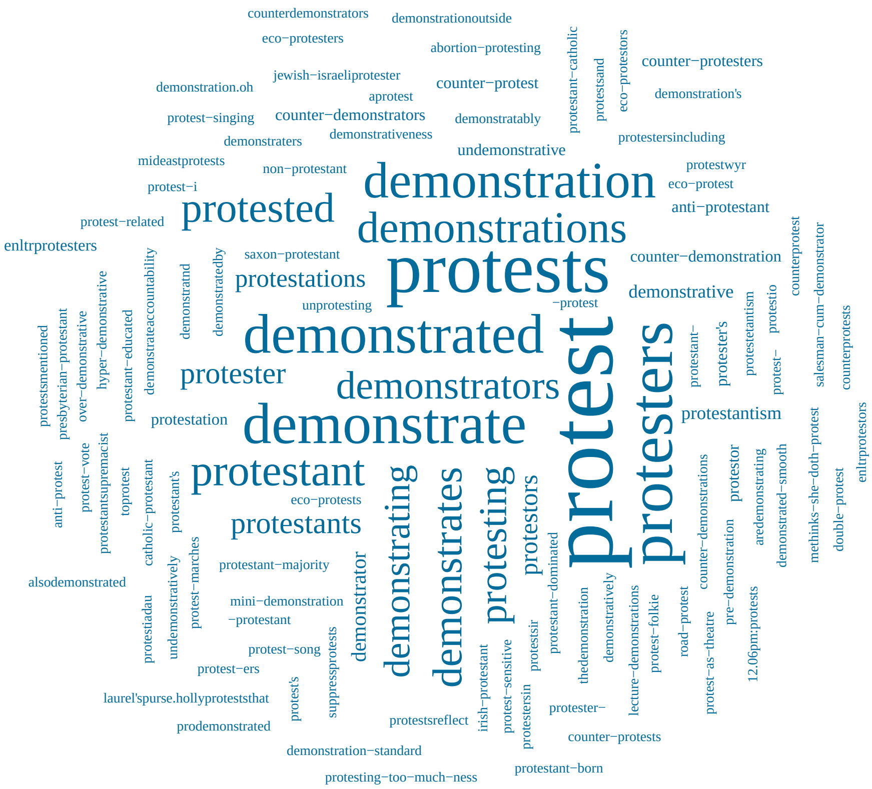


Figure 5.2: Variations of "protest" and "demonstrat". Larger Words Were Used More Often in the Pilot Articles

instead of the number of articles about protest, what grew over time is the number of superfluous data entries such as duplicates.

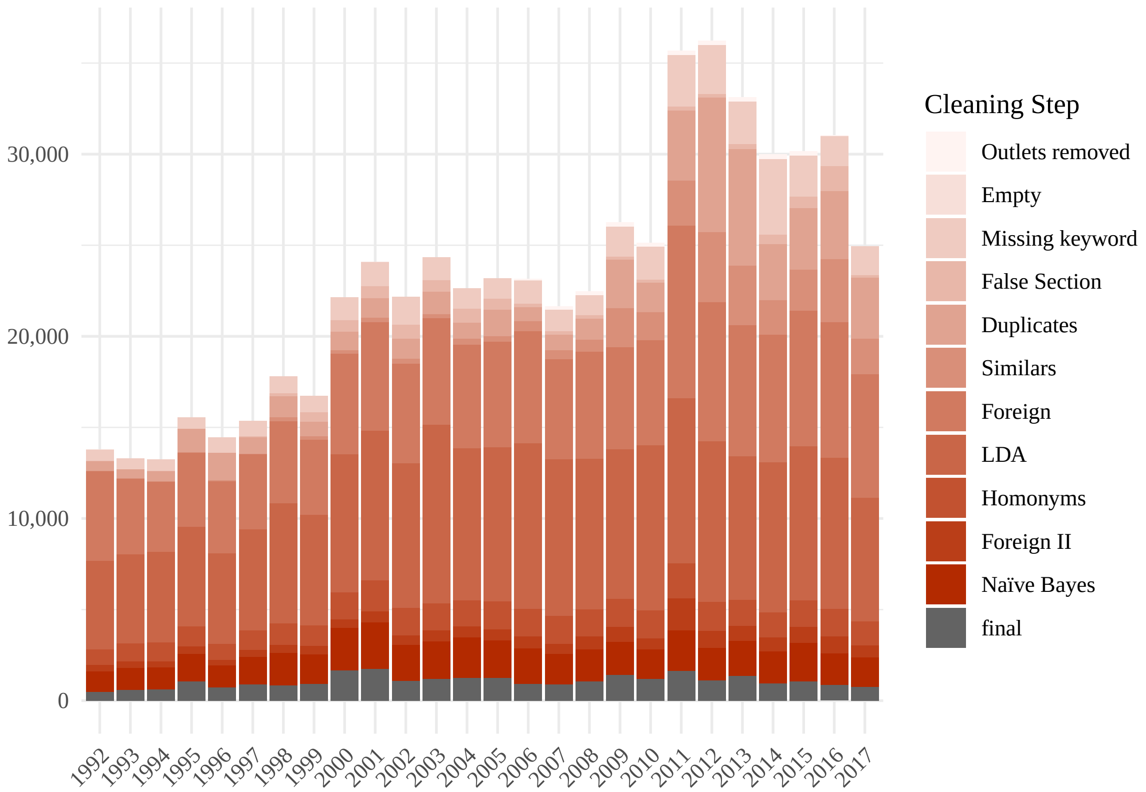


Figure 5.3: Number of Downloaded and Removed Articles per Year

The remaining section describes the different cleaning steps used to systematically yet carefully reduce the number of articles until only relevant news items remain. Table 5.2 lists the different cleaning steps along the sub-section number in which each step is explained.

5.4.1 Removing Fragments of Data Entry

Cleaning steps 1–4 removed fragments of data entry (see Table 5.2). The first step was to remove online articles, where they could be identified using the information provided by *LexisNexis*, and articles from outlets other than the eight selected newspapers. It is unclear why these articles were retrieved from *LexisNexis* in the first place, as online articles were specifically excluded and the request was only made for the eight national newspapers mentioned in Table 5.1. Yet, repeating the search consistently brought up

Table 5.2: Cleaning Steps

	Cleaning Step	Articles Remain	Articles Removed	Percent Removed	Section
0	Raw	598,962			
1	Outlets removed	596,634	2,328	0.39%	5.4.1
2	Empty	596,259	375	0.06%	5.4.1
3	Missing keyword	558,155	38,104	6.36%	5.4.1
4	False Section	549,053	9,102	1.52%	5.4.1
5	Duplicates	496,794	52,259	8.72%	5.4.2
6	Similars	470,128	26,666	4.45%	5.4.2
7	Foreign	320,237	149,891	25.03%	5.4.3
8	LDA	125,784	194,453	32.46%	5.4.4
9	Homonyms	90,226	35,558	5.94%	5.4.5
10	Foreign II	73,645	16,581	2.77%	5.4.3
11	Naïve Bayes	27,496	46,149	7.70%	5.4.6

the same articles from outlets which should have been excluded. The only plausible — yet still worrying — explanation is that since *LexisNexis* relies on manual data entry, mistakes were made during that phase. This first step removed 2,328 entries from the dataset.

The next step dealt with a similar problem. All articles which did not mention any variation of "protest" or "demonstrat" were removed ($n = 38,104$). Again, it is not clear why those articles were in the original set. However, repeating the search and looking at the results that did not directly mention a keyword revealed that the keywords occurred in the subject tags *LexisNexis* assigns to stories. Tags were not included in the downloaded data and did not appear to contain any useful information. Similarly, *LexisNexis* retrieved 375 "articles" which were entirely empty except for some meta information. It appears that *LexisNexis* uses these entries to indicate photographs not belonging to any particular article. However, as neither photographs nor captions are stored by *LexisNexis* (Deacon, 2007), these entries are entirely superfluous and were removed.

During coding of the pilot, it became clear that certain key terms, such as "FOOTBALL:" or "Golf:" — at the beginning of a headline — and "block-time updated" in the first paragraph of an article, consistently marked irrelevant articles or texts published in an irrelevant section of the newspaper. These articles were then removed as well ($n = 9,102$).

5.4.2 Removing Duplicated Entries

Cleaning steps 5–6 removed identical as well as nearly identical articles which were considered duplicated entries (see Table 5.2). As Deacon (2007) explains, *LexisNexis* data often contains duplicated entries as different editions of the same outlet are added multiple times. To get around this problem, identical articles, labelled "Duplicates" in the plot were removed ($n = 52,259$). Manual review of the dataset, however, quickly revealed more instances of seemingly identical texts. On closer inspection, it was found, that these articles often contained minor differences. Sometimes misspellings or missing punctuation were corrected. At other times, regional editions of the national outlets contained an extra sentence or two about the impact for the region.

From a theoretical view, it did not make sense though to count and analyse these almost identical articles two or more times. To remove these articles, I proceeded in two steps: first, I calculated the lexical similarity of articles. This measure indicates if two texts used the same words and is standardised between 1, meaning the exact same words were used exactly as often in two texts, and 0, meaning that there is no overlap. This measure is not entirely reliable in comparing texts with each other but is fast and computationally cheap, which makes it an excellent choice to weed out candidates which have obviously nothing in common. Text pairs with a similarity of 0.97 and above were treated as potential duplicates — a value which was chosen after manually assessing the articles at the fringes of the threshold.

In a second step, I then calculated the more accurate Levenshtein or edit distance (Levenshtein, 1966) for the potential duplicates selected in the first step using the package *stringdist* (van der Loo, 2014). The Levenshtein distance, in short, calculates how many insertions, deletions and replacements of characters are necessary to change one text into another. The relative form of this distance divides the raw result by the number of characters of the longer text, which leads to a maximum value of 1 if texts are completely different, and 0 if they are completely identical. Despite its age, calculating the edit distance is computationally extremely demanding and comparing more than a few texts on a standard computer is often not possible. Only comparing article pairs with a high lexical similarity solved this issue. This approach was later implemented into *LexisNexisTools* (Gruber, 2021) to help future research.

After again assessing articles close to various thresholds, it was determined that article pairs with a relative edit distance below 0.2 are the same news item. Where one of the two almost identical articles was published in a regional edition, the national or London edition was left in the sample and the other one removed. If both were published in a non-national edition, which one stayed in the sample was left to chance. When the same news item was published again on a different date, this was left in the dataset. In total, 26,666 articles were removed as they had an almost identical version of themselves already in the database.

5.4.3 Removing Foreign Protest Events

In cleaning Step 7 and 10, I removed articles covering non-domestic protest (see Table 5.2). Since this is a case study of protest reporting in the UK with national newspaper outlets as sources, it was important to exclude this kind of reports, as reporting about foreign protests is significantly different from reporting about domestic protest events (Boyle et al., 2012; Mueller, 1997; Shahin et al., 2016). To remove these articles, the names of foreign locations were identified through *named entity recognition*. This technique aims to identify whether a word sequence represents an entity and then classify the entity into groups such as a person, organization or location (Welbers et al., 2017). Entity recognition was performed using *spacyr* (Benoit and Matsuo, 2020). The identified locations were then matched with the GeoNames dataset (*GeoNames*, 2018) to divide locations into two categories: inside and outside the UK. Where an article did not mention at least one location within the UK and mentioned at least one foreign location, it was removed from the dataset. If an article mentioned one location in the UK or no locations at all, it was left in the dataset.

This procedure identified 149,891 articles containing only foreign locations. A manual review of a sample of 200 random to-be-removed articles revealed that in 193 (96.5%) instances the described incident happened in a foreign location. The other 7 cases either mentioned no location in the UK or no location at all. In case of the former, protagonists of the story, such as the UK prime or foreign minister, talked about a location outside the UK. In the latter cases, which were only two, location names were false positives: one article describes a quarrel in the music industry and mentioned the band Oasis, which is also a town in Mexico; in the second one the word "Branch" was picked up as being a location, even though the article did not refer to the township in Michigan but a Branch Secretary of the National Pensioners Convention. None in the sample of 200 articles contained coverage of a protest in the United Kingdom. This body of articles was therefore removed from the dataset.

Ultimately, although it performed the task, the above approach is not recommended for future studies. Recently, the R package *newsmap* (Watanabe, 2018) was published, which uses a semi-supervised Bayesian model to automatically construct a geographical dictionary. This approach is a lot faster and involves far less manual work. The package calculates scores for each country in the dictionary — the higher the score, the more likely a text speaks about a country. After some testing, I determined that the package was performing well in distinguishing between domestic and foreign protest when I used a simple decision rule: articles which have a score of 1 or above for the UK and scores below 1.2 for all other countries almost always describe domestic protest.

This was again validated using a sample of 160 randomly selected articles, which were manually coded. In 98% of the sample, the decision by the computer if an article described an event in the UK had been correct. The only caveat was that I had to leave aside predictions for Ireland as *newsmap* was not able to distinguish it from Northern Ireland. I, therefore, treated articles about Ireland as domestic until they were removed in a later cleaning step. I used *newsmap* only after the next two cleaning steps, which are discussed below. Therefore, it would have been necessary to repeat several labour intensive tasks if I wanted to replace the above-described geolocation matching strategy with *newsmap*. Since this did not seem to make sense, Figure 5.3 shows two steps which removed articles about foreign protest. Nevertheless, *newsmap* discovered 16,581 additional articles about foreign protest, which were removed from the data.

5.4.4 Removing False Positives Using Topic Models

Cleaning step 8 removed false positives using an LDA topic model (see Table 5.2). Section 4.2.2 mentioned that topic modelling, which is a form of *fully automated dimension reduction*, can extract the main themes, or *topics*, from large corpora with minimal human input. I, therefore, regard the technique as useful for finding irrelevant subsets of a corpus as it will usually find topics that are not related to the research. Articles in which these irrelevant topics are prevalent can be removed. I used common preprocessing steps: I lowercased all words, used stemming[20], I removed stopwords, punctuation, symbols, numbers, hyphens, URLs, infrequently used words (less than 6 times) and words shorter than two or longer than 80 characters with *quanteda* (Benoit, 2018). I use Latent Dirichlet Allocation (LDA) (Blei et al., 2003) as implemented in the *topicmodels* package (Grün and Hornik, 2011). The only user input necessary to run LDA is to provide the number of topics (k), similar to k-means clustering. After evaluating different k-solutions from 15 to 100 topics by assessing them on how coherent and fine-grained the models were, I decided to use 75 topics as an ideal solution.

LDA provides two sets of values: ϕ (phi), which shows for each word in the corpus how it scored on each topic. The higher the ϕ-value, the more prevalent the word in this specific topic. And θ (theta), which shows for each document in the corpus how it scored on each topic. The higher the θ-value, the more prevalent the topic is in the

[20] Stemming means to reduce different forms a word to a common base form (e.g., Manning et al., 2008). For example: car, cars, car's, cars' $\rightarrow$car.

Table 5.3: Topic Examples

	topic 3	topic 8	topic 34	topic 72
Name	Sport	Business	Protest	Football
Relevant	no	no	yes	no
1	england	compani	protest	club
2	rugbi	market	said	footbal
3	play	busi	peopl	fan
4	player	year	demonstr	leagu
5	game	sale	group	player
6	team	group	polic	manag
7	wale	share	london	game
8	tri	profit	campaign	season
9	cup	price	march	unit
10	coach	industri	organis	support
11	back	uk	yesterday	team
12	match	sell	anti	last
13	welsh	product	activist	said
14	last	trade	street	match
15	new	invest	peac	year

document. ϕ-values are used to interpret topic models as the latent dimension behind a model is otherwise unknown. Usually, the words in each topic with the highest ϕ-values are used for this. This is best explained using a few examples: Table 5.3 shows the ten words with the highest ϕ-values for a few topics (for the full list of topics, see Appendix B). These topics can be interpreted by deciding what the common theme in these topwords are. In Topic 3, for example, we see a lot of words connected to sports. Articles with a high θ-value for Topic 3 then likely talk about sports and are, hence, not important to this research.

The topics were assigned names and the table shows if they were deemed relevant. Relevance assessment was based on the words shown in Table 5.3 and the 20 articles with the highest θ-value in each topic. Overall, 29 out of 75 topics were deemed relevant. Articles in which one of the relevant topics was the first, second or third most prevalent topic, which is the topic with the highest θ-value, were left in the dataset. This removed 194,453 articles from the data and left 125,784 articles.

5.4.5 Removing Homonyms

The ninth step was to remove false positives which were added to the corpus based on additional meanings of the keywords used to download articles (see Table 5.2). While validating how many false positives were left in the dataset, it emerged that the chosen keywords were, in fact, far from unambiguous. Consider a few examples: "**protest** party", "**protest** song", "**protest** vote", "withdrew in **protest**", "cooking **demonstration**", "for **demonstration** purposes", "**demonstration** power plant" and "makeup **demonstration**". Both "protest" and "demonstration" have several *homonyms*, which means words that are written and pronounced in the same way but have a different, often unrelated, meaning (e.g., Jurafsky and Martin, 2020, pp. 493f). In other words, "protest" and "demonstration" have a number of doppelgängers that are not referring to the kind of protest this thesis studies. The homonyms can, however, often be distinguished by context.

Homonyms can potentially add a large number of false positives to the corpus, which is based on these keywords. To prevent this, I devised a strategy to remove articles based on a keyword in context analysis (kwic) performed in *quanteda* (Benoit, 2018). A kwic analysis searches for words that appear as direct neighbours or in a specified window surrounding one or several given keywords. I analyse windows of one and two words and consider the *pre*-, *post*- and *both*-context. For example, when using "protest" as keyword I get "prisoners who **protest**" as the *pre*-context; "**protest** their innocence" as the *post*-context; and "prisoners who **protest** their innocence" as the *both*-context. In this example, I consider all three contexts to be irrelevant to my research.

Where it was not immediately clear from the phrase itself, I manually assessed up to 20 example paragraphs containing a context and decided whether it was relevant. There were three kinds of phrases which were caught this way: phrases in the first category were very clearly not about protest or demonstration in a relevant sense — like the examples mentioned above. The second category consisted of phrases that describe forms of protest that are not included in the definition in Section 4.2.1: “I protest”, “s/he protested”, “protested her/his”, “would protest” and “formal protest”. In the first three examples, the protest could clearly not be called a *joint* expression as it was voiced by single actors. “would protest” indicates that no protest has actually taken place and that it is hypothetical. The last example is only used for institutional settings, for example in the cabinet or parliament, which were also excluded. In a third category, the cases were more nuanced and required more manual assessment than the others. Phrases in this category employed a specific grammatical context that is not used in regard to protest in a relevant sense. A prime example is the phrase “in protest”, “demonstrating that” and “demonstrating a”. This seemed counter-intuitive but even after extensive assessment, there was not a single case where any of those phrases were used in a story about a public protest or a demonstration.

After compiling a list of all context phrases that indicate an irrelevant usage of the keywords, I removed those phrases in the articles before running the filter process again. For example, “prisoners who *protest* their innocence” is replaced with “prisoners who ***** their innocence”.[21] When searching again for relevant keywords in these articles, they must contain at least one more “protest” or another relevant keyword to remain in the dataset. If the removed keyword was, however, the only instance, the article is removed. A second set of phrases was used to directly remove articles (see Appendix B). In total, 35,558 articles were determined to be false positives using the contexts of the keywords. These articles were removed from the database, leaving 90,226 articles.

5.4.6 Removing False Positives Using Machine Learning

The 11th and final step to remove false positives from the database was based on supervised machine learning (see Table 5.2). The reason why I performed this step last was that it seemed to require more manual labour than previous steps. However, after each cleaning step, I validated a sample and each time concluded that the number

[21] These changes were reversed before any substantial analysis was undertaken.

Table 5.4: Confusion Matrix for Validation of Naïve Bayes Model

		manual	
		0	**1**
computer	**0**	23	0
	1	15	62

metric	value
Accuracy	0.85
Precision	0.81
Recall	1.00
F1	0.89

of irrelevant articles, or false positives, was still too high to perform a meaningful analysis. Supervised machine learning was, therefore, the last resort. As explained in Section 4.2.2, the principle idea of supervised machine learning in ACA is that a model is trained based on a relatively small sample of manually coded data. After it is determined through validation that the model performs reasonably well, the remaining larger portion of data is classified by the machine.

To minimise the coding effort, I split the articles into paragraphs and coded only paragraphs which contained one of the relevant keywords. I randomly selected 1,000 paragraphs for coding and manually determined if the paragraph was talking about protest in the sense defined in Section 4.2.1. I then randomly split the coded data into training ($n = 900$) and test set ($n = 100$) to evaluate model performance. To train the model, I tested several classifiers implemented in *quanteda* (Benoit et al., 2020c). Naïve Bayes emerged as the most competent algorithm. I used the same preprocessing steps as above. A model was then trained on the training set and used to predict if the articles in the test set are relevant.

Results are shown as a confusion matrix in Table 5.4. It shows that accuracy and precision are reasonably high. Recall, on the other hand, which measures how often the model erroneously discarded relevant items, is perfect. This is an overall delightful result as it means the model creates no false negatives at all. The fact that a small portion of irrelevant articles still remains in the data is regrettable, but the dataset can now be regarded as reasonably clean.

I use the entire coded data to train another Naïve Bayes model and use it to assess which paragraphs in the remaining data are relevant. Articles remain in the dataset if they contain at least one relevant paragraph. This final step reduces the dataset by another 46,149 articles. The final database then consists of 27,496 articles, out of the original 598,962. This clean dataset is used as the basis for the analyses presented in the following chapters.

Chapter Conclusion

This chapter has described the steps to create and clean the database of media reports about protest used in this thesis. First, the selection of newspaper sites was described, based on the objective of maximising variance in the variables newspaper type and newspaper ideology. The second section then pointed out a few pitfalls when using digital news archives, specifically *LexisNexis*, but also described why the advantages of using them outweigh the disadvantages. The third section described a pilot study, aimed to mitigate one of the issues with digital news archives: as keywords are the most important tool to query all relevant articles, while retrieving as few irrelevant articles as possible, selecting the right keywords is crucial. Most importantly, omitting important keywords means relevant news items would be missing from the research. Therefore, a procedure was developed to ensure that all relevant articles are found by downloading a sample with all potential keywords that are even slightly related to protest. In the end though, the two most and only important keywords turned out to be “protest” and “demonstration” plus a few variations and (mis)spellings of these words.

The remaining section described how I cleaned the dataset from false positives after I had downloaded articles from *LexisNexis*. Including irrelevant articles in the dataset would likely bias analysis results. Usually, articles are removed during coding as coders notice that an article does not speak to the topic at hand. Yet, since the dataset for this thesis is vast, not all articles could be manually assessed. The step to clean the dataset was thus especially crucial. Like with the pilot, however, hindsight provided the conclusion that not all cleaning steps had ultimately been necessary — although all were successful in themselves. Specifically, removing articles on foreign protest based on a dictionary of locations was outperformed by the algorithm implemented in the *newsmap* R package (Watanabe, 2018), which also used far fewer resources. The strategy to employ topic models likewise required far more manual work than expected, while the resulting database was still unfit for further analysis. To be clear, these steps performed reasonably well in removing redundant data. However, neither cleaned the sample to a point where the planned analysis could have been conducted. Instead, using a Naïve Bayes classifier, in combination with the *newsmap* package, would probably have sufficed to arrive at the same reasonably clean dataset. On the other hand, this was hard to anticipate beforehand, and both steps seemed good choices to bypass the labour-intensive manual coding needed for machine learning classification. Ultimately, I hope that some of the described work is useful for others facing a similar task.

The goal when constructing the database was that it should contain all articles about protest published in the eight selected national UK newspaper in the selected time frame. However, as most of the articles in the data are only ever coded by the computer, it is especially important to make sure that as few irrelevant articles as possible are still present in the final dataset. The drastic reduction from originally 598,962 downloaded news items to 27,496 newspaper articles in the final sample testifies to the challenge this represented.

Chapter 6

Framing Protest

This chapter systematically explores how protest was framed by national UK mainstream newspapers since 1992. More specifically, the purpose of this chapter is to answer two of the three research questions: *How do British newspapers frame the coverage of domestic protest events?* And *how — if at all — did the framing of protest reporting change over the last 26 years?* As well as the respective hypotheses, H1 (*Stories on protest events mainly use delegitimising framing as described in the protest paradigm literature*), H2a (*Delegitimising framing decreases in salience over time*), and H2b (*Legitimising framing increases in salience over time*). To do this, I code the large corpus of news reports on UK-based protests developed in Chapter 5 ($N = 27,496$) using the frame analysis method outlined in Section 4.2.3. The chapter is divided into three parts.

Firstly, Section 6.1 discusses the manual coding effort outlined in steps 1 to 3 of the approach (see Figure 4.1). Specifically, I developed a codebook based on the idea that frames consist of smaller units, that is frame elements. A random sample of 500 newspaper articles was then coded. Intercoder reliability was found to be at a high level, demonstrating the advantages of the approach used.

The second section focuses on the next steps, whose overall aim is to *identify frames*, which means to inductively and systematically determine what the main frames in the discussion are. This is done inductively from the material via dimension reduction (see Figure 4.1, Step 4 and 5). Key to this approach is that frames are understood as recurring usage patterns of the same frame elements (see Section 4.2.3). Following Matthes and Kohring (2008) and David et al. (2011), I first use cluster analysis to group similar articles based on the codes they have been assigned. As frames are seen here as recurring pattern of the same frame elements, the distinct features of each frame become visible as cluster means of codes. As cluster analysis groups together articles with similar codes assigned to them, the emerging groups should consist of articles using the same frames. By and large, the emerging groups appear reasonably meaningful. Additionally, I use factor analysis, which resulted in a similar, yet more nuanced set of frames. Whereas clustering attempts to find the best grouping for cases, factor analysis groups variables directly. Doing so, it allows mixed membership of cases, meaning that each article may contain several frames, which makes theoretically more sense. The direct comparison of the two methods showed that factor analysis should be regarded as the theoretical and empirically better suited technique for identifying frames through dimension reduction.

In the next section, I explain how the *identified* frames were then used to *code* newspaper articles, meaning that it was established whether each article used a frame or not (see steps 6 to 8 in Figure 4.1). This is initially done for the training sample — the 500 manually coded articles — using factor scores from the factor analysis. With this step completed, it is possible to train supervised machine learning (SML) algorithms to determine if a frame is present in an article based on the vocabulary used in it. SML models were validated by classifying a held-out portion of the training set and comparing the classes assigned by the machine with the "real" classes assigned through manual coding. After validity of the models was established, the remaining data is automatically coded. To reiterate, frames are first inductively *identified*, then *coded* in a manually coded training set, before this set is used to train SML models, which code frames in the remaining data.

The final section discusses the results. Specifically, seven main frames were *identified* in the mainstream news media coverage of domestic protest events. Four of the these are linked to the *protest paradigm*, meaning they focus on the trouble and nuisance caused by protests and usually portray protesters in a negative light. Two of the seven frames, however, are more legitimising for protesters' message. One frame highlights police brutality and is neither legitimising nor delegitimising. Once these frames were *coded* in all articles, it emerged that delegitimising frames were overall more salient in articles about protest than legitimising coverage — confirming H1. In terms of how *patterns of protest reporting change over the last 26 years* I found that a stable majority of articles uses delegitimising frames throughout the entire time frame. At the same time, however, a substantial and growing number of articles employ legitimising frames. This means that H2a is rejected while H2b is validated.

6.1 Manual Coding

6.1.1 Frame Elements and Codebook

The first step of the *frame identification* procedure outlined in Section 4.2.3 is to develop a codebook for the manual content analysis. Since frames are an implicit construct, coding them holistically can decrease reliability and validity as both frame identification and document classification involve a high degree of subjective judgement (Matthes and Kohring, 2008). As explained in Section 4.2.3, frames are understood here to be different combinations of certain frame elements. To find these combinations and, therefore, the employed frames, I proceeded in two steps: frame elements were manually coded separately before identifying the main frames, using dimension reduction techniques. This procedure relies on smaller, more transparent human decisions as frame elements are far more explicit and easier to code than holistic frames. The codebook was constructed using a combination of inductively and deductively discovered codes. This way, it was possible to draw on the body of literature on the topic reviewed in Chapter 2, while new categories found in the dataset, which covers a longer time frame and broader scope than previous research, could also be seamlessly included.

In theory, there are a number of different possible ways to deconstruct a frame into elements and subsequently content analysis variables. van Gorp (2007), for example, introduces three special elements that can be combined to a "frame package": "the manifest framing devices, the manifest or latent reasoning devices, and an implicit cultural phenomenon that displays the package as a whole" (van Gorp, 2007, p. 64). These elements can have several sub-categories that vary in number and depth of the hierarchy.

Matthes and Kohring (2008) suggest that the four elements of a frame in Entman's (1993) definition should form the first layer of a coding scheme. Framing, according to this definition "promote[s] a particular *problem definition*, *causal interpretation*, *moral evaluation*, and/or *treatment recommendation* for the item described" (Entman, 1993, p. 52, emphasis added). Employing this definition has two advantages: firstly, Entman's definition is the most widely accepted definition of framing in a currently rather fractured field of research — framing analysis. Secondly, it makes the mechanisms behind framing more explicit than other definitions by offering four distinct elements to work with. However, *problem definition*, *causal interpretation*, *moral evaluation*, and *treatment recommendation* themselves turn out to be still relatively abstract categories. They are, therefore, further divided into content analysis variables following the explanation of each element by Entman (1993) and suggestions by Matthes and Kohring (2008) on how to improve content analysis of media frames (see the column *Variable* in Table 6.1).[22]

[22] More details about the specific variables can be found in Appendix A.

Table 6.1: Codebook for Manual Content Analysis

Frame Element	Variable	Code	Description
Problem Definition	Topic	Clash	Confrontation with the police, not necessarily violent.
		Violence/Crime	Violence, vandalism and destruction of public or private property surrounding a protest.
		Cause	The cause of why a protest took place or the goals of the protesters.
		Protesters	The appearance, mental ability, visual deviance and oddities of the protesters (including pathologising their protest group, social movement or subculture) or their 'real' underlying motives (e.g., adolescent anger, lust for destruction, chaos and anarchy or astroturfing allegations).
		Spectacle	Highlighting the entertaining or spectacle aspects (e.g., stunts or performances by protesters or the presence of celebrities).
		Protest tactics*	Interrogation of protesters' tactics and why they were adopted.
		Policing tactics*	Discussion about how the police, security forces or laws should deal with protesters.
		Response*	Repeating or discussing the response to the protester's demands (e.g., the prime minister's response to demands to end a policy).
		Confrontation/Showdown	The protests are part of a confrontation between two groups (e.g., between political parties). Usually, a 'showdown' rhetoric is applied.
		Public opinion	The protest represents a minority/majority of public opinion (operationalised through polls, interviews or reference to norms).
		Nuisance	The protest caused inconvenience to regular citizens and the government.
		Other activism*	Other actions by the same group who initiated protest (like letters or action in courts).
		Judicial prosecution*	Prosecutionof protest-related actions in court (e.g., court case about violence of protesters or police).
		Effect of protest*	Discussion about the effect a protest had or lack thereof (e.g., starting a public debate about a topic, leading to policy changes, etc.).
		Event	Description of the event (e.g., size, marching route or what protesters did) but not highlighting the entertaining or spectacle aspects (see Topic Spectacle).
		Media*†	Scrutiny of the attention the media spent on a protest.
	Actor	Protesters	The people engaging in a protest.
		Police	Members of the police force engaging with protesters.
		Officials	Representatives of government organisations.
		Business	Representatives of business organisations.
		Other Political Elite*	Other political players such as members of the opposition party or regional parliaments.
		Other*	Other non-elite actors (e.g., motorists, local residents, counter-demonstrators or unidentified people).

* Category found inductively from the coded material instead of the literature.

† Category later removed due to infrequent use.

Table 6.1: Codebook for Manual Content Analysis (continued)

Frame Element	Variable	Code	Description
Causal Attribution	**Benefit Attribution**	Police	Police is responsible for the benefit (e.g., by reinstating order).
		Protesters	Protesters are responsible for a benefit (e.g., by being entertaining).
		Officials	Officials are responsible for a benefit (e.g., by preventing chaos protesters try to inflict).
	Risk Attribution	Protesters	Protesters are responsible for risk (e.g., vandalism, attacks on police or nuisance).
		Police	Police are responsible for risk (e.g., unnecessary clashes with peaceful protesters).
		Business*	Business actors are responsible for risk.
		Officials*	Officials are responsible for risk (e.g., by signing a bad law).
		Media*	The media are responsible for risk (e.g., problematic reporting).
		Other Pol. Elite*	Other political actors such as members of the opposition party or regional parliaments are responsible for a risk.
		Other*	Other actors are responsible for a risk, often by attacking protesters (e.g., motorists, local residents, counter-demonstrators or unidentified people).
Moral Evaluation	**Benefit**	Reinstating public order	Usually attributed to police who reinstate public order after it was strained by the protest.
		(Just) Cause*	Usually attributed to protesters when their actions are seen as part of a struggle for a good cause.
		Initiated public debate*	Usually attributed to protesters when their actions have caused a (necessary) public debate.

* Category found inductively from the coded material instead of the literature.

† Category later removed due to infrequent use.

Table 6.1: Codebook for Manual Content Analysis (continued)

Frame Element	Variable	Code	Description
Moral Evaluation	Risk	Grievance*	The grievance which the protest is addressing is the risk (e.g., if the protest was against the tuition fees, the grievance could be that working-class students will not be able to afford education).
		Public safety	The protest is a risk for public safety.
		Property destruction	Property was destroyed due to protest.
		Harassment*	Protesters are harassing people instead of engaging in arguments (e.g., screaming at clients in family planning facilities instead of campaigning for anti-abortion rules or intimidating workers of animal testing facilities instead of campaigning for stricter animal welfare laws).
		Breaking laws*	Breaking the law but neither destroying anything nor harming anyone (e.g., entering parliament without permission).
		Decay of morals or other social norms	The protesters are a fringe group of freaks who set a morally bad example for others and disturb the general political consensus.
		Nuisance	Protest is bothersome to the daily lives of citizens and the government yet impotent as a political tool.
		Harm discussion*	The protest or actions of protesters harm the discussion about a topic (e.g., by shouting down opponents or by distracting from the cause).
		Trivializing (political) discussion	The protesters spoil serious (political) discussions with their childish, insane or uninformed arguments or false claims.
		Bad for business*	Protests have a negative impact on business revenues.
		Costs of demonstrations	The costs of a demonstration (e.g., clear up and police deployment costs of demonstrations) including what this might cause (e.g., burden public budget or spread police thin for other tasks).
		Suppression/Censorship	Usually attributed to police or officials who allegedly try to undermine protest or make it impossible.
Treatment	Judgement	Negative	The protests were/are bad (e.g., protesters should go home).
		Positive	The protests were/are good (e.g., protests made aware of a problem).

* Category found inductively from the coded material instead of the literature.

† Category later removed due to infrequent use.

Why were the frame elements divided into codes in this specific way? ***Problem Definition*** can be conceived by coding the topic, or central issue that is discussed, and the main actor described in a text. Both topic and actor can occur only once, as codes of the same variable were assigned to each paragraph mutually exclusive.[23] As an example, the main problem discussed in a text about a protest could be the violence surrounding an event — either the violence against people or property. The main actor, or rather actors, could be the protesters if they are the ones described as resorting to violence, or, it could be the police or officials who try to stop the violence.[24] Taken together it becomes clear that violence and protesters or violence and police form two distinct problem definitions, although it is not yet clear who is responsible for the problem.

Causal attribution and ***Moral evaluation*** — the second and third frame elements — are closely linked to each other. *Causal attribution* is conceptualised to contain an attribution of who is responsible for a *benefit* or a *risk* or both. *Moral evaluation* entails said benefits and risks. In the example above, a text could portray protesters as being responsible for the risk that reported violence caused. They do not necessarily need to be the main actor in the text for that. For example, the risk could be that public safety was endangered or property was destroyed. The police could be portrayed as the actor and as being responsible for the benefit of ending the violence by detaining violent protesters. However, a news report could also flip the attribution roles: actions risking public safety could be attributed to the police if they started to clash with protesters unprovoked and without apparent necessity. The protesters could be portrayed to cause a benefit, for example, if their message is portrayed as important and they are seen as demonstrating for a good cause. Depending on the benefit and risk and who is deemed responsible for that, the example above — *violence* being the topic and *protesters* being the actor — can be modified substantially: protesters might be the main actor in a story about a peaceful protest that turned into violence, yet the risk

[23] More details about coding paragraphs with more than one topic/actor are discussed in Section 6.1.2.

[24] It is also often described that the police started the violence. The example is just an illustration.

of endangering public safety might be attributed to the police. Again, each paragraph will be coded only for one *risk* or *benefit* and *risk/benefit attribution*. Yet an article might discuss several *risks/benefits* or make different actors responsible for the same *risk/benefit* over several paragraphs.

Entman's fourth frame element, *treatment recommendation*, is conceptualised to contain either positive or negative ***judgement***. In the case of protests, this usually means that if a protest is judged positively, the treatment recommendation is that protests should continue until a stated goal is reached. If the *judgement* is negative, the treatment recommendation is that protesters should disperse and go home or even be persecuted for their actions. The treatment recommendation can be explicit, but usually it is not. When it comes to protest in the UK, there is apparently a broad consensus that explicitly arguing for a protest to stop would be out of line for a journalist. This does not mean, however, that news articles shy away from implying positive or negative judgement of an event through a combination of other frame elements. When the *topic* in all or most paragraphs of an article is *violence*, for example, and all resulting *risks* are attributed to *protesters* while neither reasons nor arguments of the protest are reported, the resulting frame would reflect the negative slant, even if negative judgement was not directly spelled out.

For each of the variables, there are a number of codes. As explained in Section 4.3.1 and shown in Figure 4.1, there are two sources for codes: the literature and the coded material. In Table 6.1, codes that originated from the material are marked with an asterisk, the other ones were derived from the literature. Usually these codes were adapted to fit the coding scheme. In McLeod and Hertog (1999), for example, a *violent crime story* frame is described. From this, I derive the *problem definition topic: crime* and *actors protesters* and *police*. In terms of *moral evaluation,* the *risks* of a diminished *public safety* and *destroyed property* were added to the codebook as well as the *benefit* of *reinstating public order*. For the *causal interpretation*, both *risks* are attributed to the *protesters*, while the *benefit* is attributed to the *police*. The *treatment recommendation* is not explicit, although a negative judgement is implied through the other elements.

Note that this means that frames are conceptualised in a four-level hierarchy: frames are divided into frame elements, which are divided into variables, which are divided into codes. Since the frame level is unknown before dimension reduction, Table 6.1 only shows the other three levels. Since the remaining chapter will discuss the codes in Table 6.1 repeatedly, the following notation is used from here on: when discussing the code *cause* in the variable *topic* in the frame element *problem definition*, I will use *topic: cause* (i.e., *variable: code*) in lowercase and italic font — the description of the frame element is omitted. When discussing the code *grievance* in the variable *risk* in the frame element *moral evaluation*, I will use *risk: grievance* and so forth.

Codes with an asterisk in Table 6.1 were not used in any previous studies, but were discovered in the database used in this project, which is large and comprehensive and covers a range of different types of protest. One example for this is the *topic: judicial persecution*: several articles in the datasat contained detailed descriptions of how protesters defended themselves in court after they were arrested during a protest or the sentence they received. Prior research of protest coverage does not mention trials as these usually take place months after an event and hence outside the analysed time frames. The newly added codes can be considered a strength of the chosen approach. Despite a vast body of literature, occasionally variables emerged during coding that were not previously covered. Combining inductive and deductive techniques, therefore, appears to be the right choice.

6.1.2 Coding Procedure

For the first step of the coding process, 500 randomly chosen articles were coded using the codes presented in Table 6.1. As mentioned in Section 4.2, a goal of the procedure was to make coding decisions as small and explicit as possible to increase validity and reliability. For that reason, articles were split into paragraphs before coding, even though the unit for the framing analysis is the article. The coding task was defined as deciding which — if any — *problem definition*, *causal interpretation*, *moral evaluation*, and *treatment recommendation* were used in a paragraph. To reiterate, the reason for choosing the paragraph as the unit for manual content analysis was that coding frame elements for the entire article involves demanding, and thus potentially rather subjective, judgements for the coder while the sentence level usually does not contain enough information to make any decision at all. This turned out to be the right choice during coding as many articles, especially longer ones, often switch perspectives between paragraphs. Articles tend to highlight what one actor has done or said in one part of the story and focus on another party in a later part. Coding an entire article at once, therefore, appears demanding if not impossible. A paragraph, by comparison, usually comprises only one argument or thought. While this problem is not completely avoided by coding on the paragraph level, decisions are usually more straightforward and hence more reliable.

This, however, lead to a different problem: it is not that common for individual paragraphs to contain all frame elements at once. Theoretically speaking, this is no problem for the applied method since not every frame necessarily contains all elements. The absence of, for example, a risk might be a distinct feature of a frame that portrays a story decidedly positive (Entman, 1993; Matthes and Kohring, 2008). Practically speaking though, parts of the same frame were often spread over multiple paragraphs when one would mention a topic and an actor, setting the stage to describe moral evaluation and causal attribution later on in the text. Codes were, therefore, combined on the article level in a later step — more on that in Section 6.2. Again, the choice to code on the paragraph level should thus be seen as an aid to code more systematically, while the framing is later determined on the article level.

Paragraphs were coded in order of appearance and codes within the same variable were assigned mutually exclusive within a paragraph. Where, for example, two topics or actors were present in a paragraph, the one that was featured more prominently was chosen. If two or more codes are discussed equally prominent, the one that is more specific is chosen. To reflect this ranking, Table 6.1 is ordered from more specific to less specific within each variable.[25]

Assigning individual explicit variables from the exhaustive codebook proved to make coding relatively easy compared to more implicit categories, such as holistic frames or story types. Progressing through the material it became clear that the variables and categories included in the codebook were sufficient to map the coverage of protest exhaustively, even though the sample contains a wide variety of reports on starkly different protest events. To be clear though: codes in the codebook are not meant to be exhaustive in the sense that they can be used to describe all text passages in all of the selected articles. Only text passages that had something to do with protest were coded and the codebook reflects this decision. In paragraphs where neither the protest nor grievances of the protest were mentioned, no code was assigned.

6.1.3 Inter-Coder Reliability

For this thesis, the reliability data consists of a random sample of 600 paragraphs (approx. 20% of the entire sample that was manually coded) which were initially coded by myself. This sample was also coded by two other PhD researchers, one from Germany and one from Spain, who were assigned 300 paragraphs each. They were provided with: the codebook in Table 6.1; an extended version of the codebook which includes a brief explanation of the approach, the goal of the research, more detailed descriptions of the codes than provided in Table 6.1 as well as several examples for each code from the data[26]; and an appropriate level of training, including an initial phase in which they coded several articles under supervision, which were not included in the reliability data.

[25] For details on coding rules see Appendix A.

[26] The extended codebook is included in Appendix A.

The coding process technically involves two interlinked but separate tasks: 1. identifying for each paragraph if a topic, actor, benefit attribution, risk attribution, benefit, risk and/or judgement are present (*identification*); and 2. to classify which codes describe the paragraph best (*classification*). Agreement for those two tasks was, therefore, measured separately.[27] Besides simple pairwise agreement between coders, I report Cohen's κ (kappa) (Cohen, 1960) and Krippendorff's α (alpha) (Krippendorff, 2004)[28] for both tasks. This is done to correct the impact of chance, as described in Section 4.2.4. In the case here, κ and α are nearly identical, as the advantage of Krippendorff's α is only that it is more flexible and can account for different sample sizes, missing data and more than two coders, which are no problem under the current setting anyway (Krippendorff, 2004, pp. 244-250). It is nevertheless worthwhile to include both, as they are usually interpreted differently.

Specifically, while there are no commonly accepted thresholds at which Cohen's κ and Krippendorff's α are accepted as reliable, the following interpretations are often cited: $\kappa < 0.00$ is considered poor agreement, $\kappa = 0.00$-0.20 slight agreement, $\kappa = 0.21$-0.40 fair agreement, $\kappa = 0.61$-0.80 substantial agreement and κ values of 0.81 and above are considered to represent almost perfect agreement between coders (Landis and Koch, 1977). For the α-coefficient, Krippendorff suggests that researchers can rely — in most applications — on α-levels greater than 0.8 and should consider reliabilities between $\alpha = 0.667$ and $\alpha = 0.8$ only for drawing tentative conclusions (Krippendorff, 2004, pp. 241-243). This means that there is disagreement between interpretations of κ and α levels between 0.667 and 0.8, which are accepted as substantial agreement when κ is calculated but are seen more sceptically when one uses the α-agreement measure.

As can be seen in Table 6.2, which lists results for the *identification* coding, Cohen's κ (kappa) and Krippendorff's α (alpha) for several of the variable range in this prob-

[27] Measuring reliability for the two tasks separately must be considered more lenient than the alternative which would be to treat the coding process as one task, treating non-identification as disagreement in the classification. This seems unnecessarily strict in this case, however, since aggregation of the codes on the article level usually levels small differences while coding individual paragraphs.

[28] Both calculated using the R package *irr* (Gamer et al., 2019).

Table 6.2: Inter-Coder Reliability for Identification Task

	Cohen's Kappa	Krippendorff's Alpha	Agreement
Topic	0.7267	0.7267	0.9799
Actor	0.7062	0.7064	0.8594
Benefit Attribution	0.7981	0.7982	0.9960
Risk Attribution	0.8378	0.8378	0.9598
Benefit	0.7462	0.7462	0.9920
Risk	0.7571	0.7570	0.8996
Judgement	0.8360	0.8360	0.9880

Table 6.3: Inter-Coder Reliability for Classification Task

	Cohen's Kappa	Krippendorff's Alpha	Agreement
Topic	0.8453	0.8458	0.8661
Actor	0.9433	0.9437	0.9710
Benefit Attribution	1.0000	1.0000	1.0000
Risk Attribution	1.0000	1.0000	1.0000
Benefit	0.9101	0.9125	0.9412
Risk	0.8752	0.8766	0.9000
Judgement Positive	1.0000	1.0000	1.0000

lematic band: all variables except *risk attribution* and *judgement* range above 0.667 but below 0.8. Agreement ranges between 0.71 (Actor) and 0.84 (Risk Attribution) for both Cohen's κ (kappa) and Krippendorff's α (alpha). Following Krippendorff's (2004, pp. 241-243) guidelines, there is, therefore, some reason for concern about reliability of the data for the *identification* task.

In contrast, interpretations of the metrics for the *classification* task are unequivocally good: Table 6.3 shows that agreement ranges between 0.85 (Topic) and a perfect value of 1.00 (Benefit Attribution, Risk Attribution, Judgement Positive) for both Cohen's κ and Krippendorff's α. This means that agreement between coders is perfect or almost perfect for all variables.

The picture that thus emerges is that the data can be considered reliable when it comes to the *classification* task, while Krippendorff's (2004) interpretation of his coefficient calls for caution in case of *identification*. It seems that providing coders with the option to not code a variable at all might be a potential flaw of the applied procedure, the specific codebook or the material randomly selected for coding. Yet I consider it

Table 6.4: Removed and Merged Rare Codes

Code	Merged With
topic: media	removed
benefit: initiated public debate	benefit: (just) cause
benefit: initiated public debate	benefit: (just) cause
risk attribution: business	risk attribution: officials
risk attribution: media	risk attribution: other
actor: bystander	actor: other
risk attribution: other pol. elite	risk attribution: officials
benefit attribution: officials	benefit attribution: police

unlikely that this biases the conclusions of the study. It should also be noted that for a binary classification task with unbalanced classes, Krippendorff's α measure tends to overemphasize occasional disagreement (Burscher et al., 2014). This is the reason why κ and α are relatively low for the variables *topic*, *benefit attribution*, *benefit* and *judgement* in Table 6.2, while the pairwise agreement is 98% or above.

Coding the 500 random articles resulted in 5044 codes over 2815 paragraphs. Figure 6.1 shows how often each code was assigned in the sample. It shows that some codes were extremely rare. Since this might lead to problems during the statistical analysis of the data, I decided to exclude codes which do not appear at least 20 times in the material. Table 6.4 shows which codes were removed or merged with other codes.

6.2 Frame Identification

After manual coding, the next steps were to determine frames through dimension reduction techniques (*frame identification*) and code which frame was used in each article (*frame coding*). The idea here is that each text now has a value on 42 dimensions (i.e., the codes used during manual coding), which is a detailed description of each text but not practical in terms of answering the question of how protest was portrayed (see the example in Table 6.5). Instead, the aim is to use this data to reveal frames by determining which frame elements are often used together in texts (see Figure 4.1, Step 4). To do this, I use two different techniques: first, as suggested by Matthes and Kohring (2008) and David et al. (2011), I use cluster analysis to group articles together

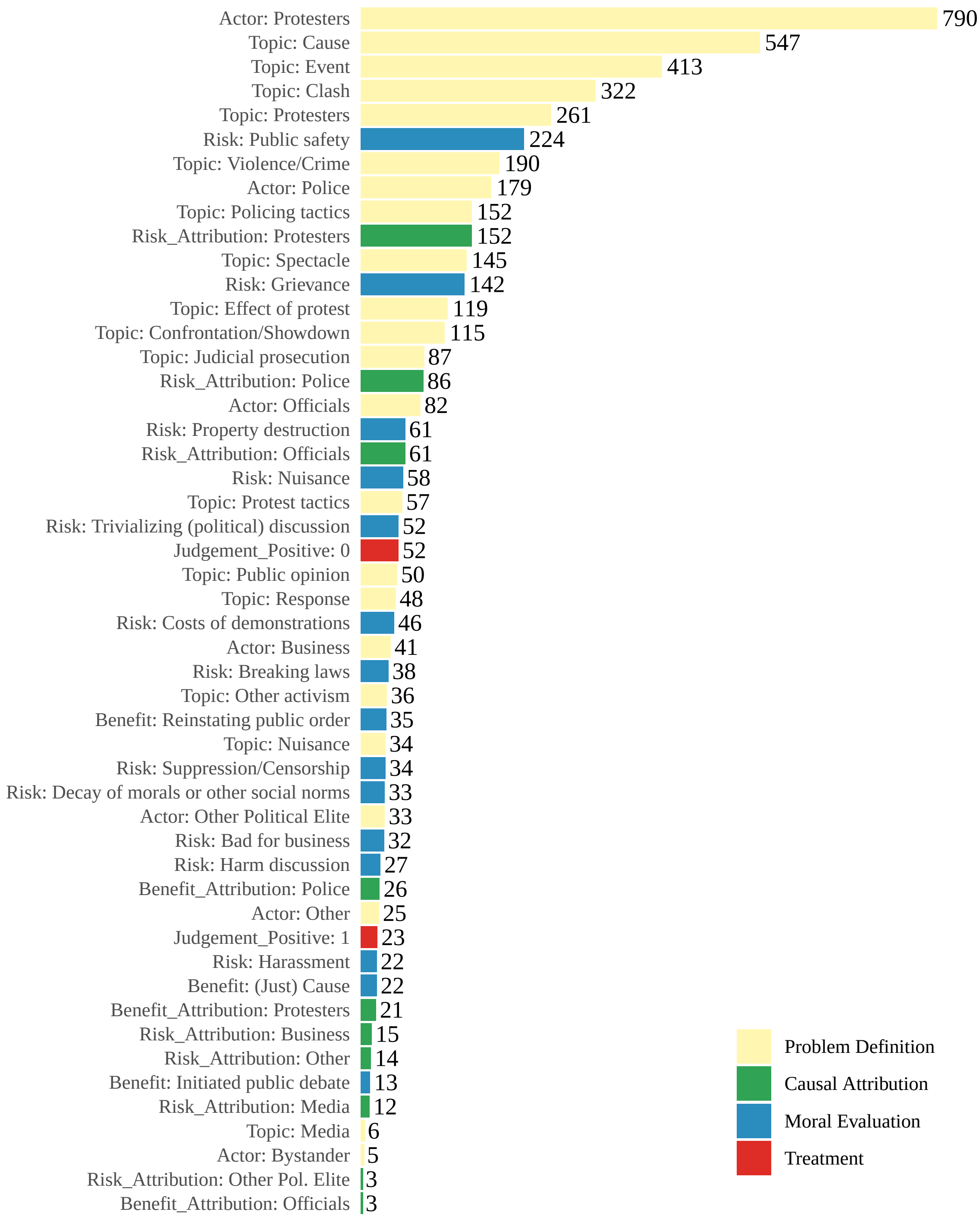


Figure 6.1: Number of Times Each Code Was Assigned

which are similar in terms of what frame elements they employ. With this procedure, the frame description emerges from cluster means, as they represent which codes were used more or less often for articles in each cluster. In a second step, I perform an additional factor analysis. In contrast to cluster analysis, which assesses similarity between cases, factor analysis reduces dimensionality of the data by assessing which of the variables belong to the same underlying latent dimension. Dimension reduction reduces the complexity of interpreting the data. Again the idea is that the remaining, reduced dimensions represent the commonly employed frames in the data.

6.2.1 Data Structure

In Table 6.5 the data structure which was used for both cluster and factor analysis is exemplified. The actual data has 42 columns (one for each variable) and 500 rows (one for each coded article). As mentioned above, I decided to code on the paragraph level. In a second step, the codes on the paragraph level are aggregated to article means before further analysis, resulting in the data structure demonstrated in Table 6.5. For example, if *violence/crime* was assigned as the *topic* in five paragraphs and *event* in the remaining five of a ten paragraph long article, the value for *topic* would be 0.5 for each *violence/crime* and *event* and 0 for all remaining possible codes in *topic*. The article mean is a suitable measure here as it normalises the value between articles and variables. That means that even if an article has many paragraphs and another one only has few, the values for each variable will lie between 1 and 0 in both cases. The same is true for often versus rarely occurring codes (Hennig et al., 2016, pp. 716-717). Additionally, the article means preserve most of the detail gained from coding the smaller unit. It could be argued that when coding on the article level, coders (subjectively) decide about the most important *topic*, *actor, risk* and so forth, even though more than one might be found per category. This could be emulated by choosing only one code per variable, namely the one which reaches a mean above 0.5.

Table 6.5: Example Matrix of Coded Articles

Art_ID	Topic: Cause	Topic: Violence/Crime	⋮	Actor: Police	Actor: Protesters	⋮	Risk: Grievance	Risk: Public safety	Risk: Property destruction	⋮	Risk_Attribution: Officials	Risk_Attribution: Police	Risk_Attribution: Protesters	⋮	Judgement: Negative	Judgement: Positive
1	0.00	0.83	...	0.83	0.17	...	0.00	1.00	0	...	0.00	0	1.00	...	1	0
2	0.80	0.00	...	0.00	1.00	...	1.00	0.00	0	...	1.00	0	0.00	...	0	0
3	0.83	0.00	...	0.00	0.00	...	0.83	0.00	0	...	1.00	0	0.00	...	0	1
4	0.00	1.00	...	0.00	1.00	...	0.00	1.00	0	...	0.00	0	1.00	...	0	0
5	0.67	0.00	...	0.00	0.71	...	0.83	0.17	0	...	0.67	0	0.33	...	0	1

* '...' represents omitted columns.

Yet, the information if an article does try to balance between different views would be lost in that way.

6.2.2 Cluster Analysis

Cluster analysis is a statistical procedure in which multivariate data, in this case codes assigned to each document, are grouped together so that cases similar to each other belong in one group and cases dissimilar to each other belong to different groups (e.g., King, 2015). It has been widely used in political science whenever there are too many variables in the data that make it difficult for a human to reliably sort cases, such as political systems, countries or parties, into groups (König, 2018). Following Matthes and Kohring (2008) and David et al. (2011), I use this method to group articles. To recap, articles are seen as using the same frame if they employ a similar combination of frame elements. Using cluster analysis on the codes assigned to the material, therefore, seems appropriate to identify which articles use the same frames. By grouping together similar articles, recurring patterns emerge by comparing which codes appear more often in one of the clusters. These empirically determined patterns become the defining features of the *identified* frames.

To perform the clustering, I chose the Hartigan-Wong k-means algorithm implemented in R (R Core Team, 2021). K-means continues to be one the most popular approaches for partitioning datasets (Mirkin, 2016, p. 34). The reasons are that the algorithm is easy to use, understand and interpret, while being very flexible and still holding up to much newer approaches in terms of performance (Everitt, 2011, p. 123). One problem with the k-means approach is that the result might be affected by the random choice of the starting partition which might lead to different optima from run to run (Everitt, 2011, 122). To eliminate this problem, I started the algorithm 100 times in R before selecting the solution with the lowest within-class sum of squares.

A crucial step in k-means cluster analysis is to decide on an optimal number of starting centres and, hence, clusters (k). However, there is no consensus about what the best approach to find k is, as this varies significantly between datasets.[29] Often approaches are, therefore, combined (Hennig et al., 2016, pp. 608-611). I use the procedure suggested by Charrad et al. (2014) who take this idea one step further by simply combining *all* available indices in their systematic approach. To be specific, they use the whole list of available indices found by Milligan and Cooper (1985) and extend it by newer approaches to comprise a total of 30 different methods. To calculate the indices, k-means clustering is performed once for each theoretically sensible k and each index is calculated for each solution. The overall optimal number of clusters is determined to be the one on which a majority of the indices agree. Charrad et al. (2014) show that selecting the k on which most indices agree outperforms any individual index.

Figure 6.2 plots the results from different indices for solutions with 2 to 15 clusters — which was the minimum and maximum thought to make theoretical sense. The three-cluster solution is suggested by 9 indices, a 2-cluster and 12-cluster solution is suggested by four indices, a 15-cluster solution by 3 and all other solutions are suggested by 2 indices or less. To test the robustness of this result, the data was

[29] Additionally, it is also disputed what "best" means as usually several legitimate clusterings exist in a dataset.

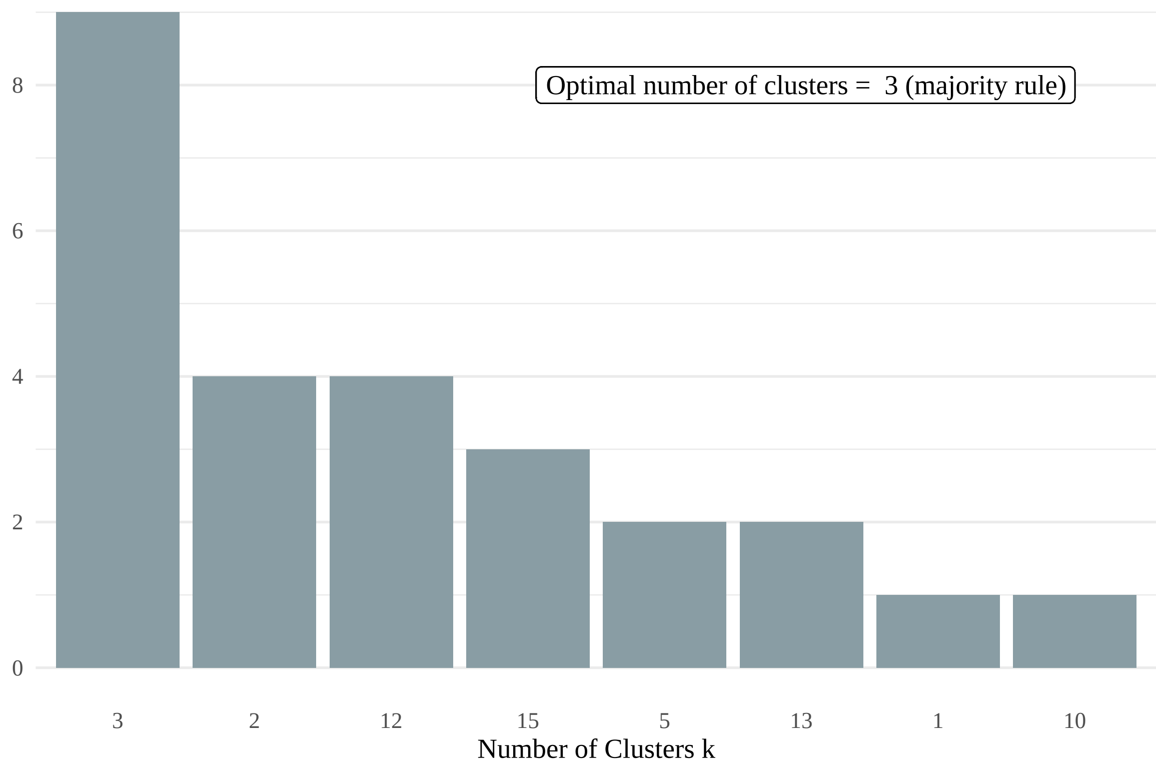


Figure 6.2: Frequency Among All Indices for Best Cluster Solutions

randomly divided into two subsets and the indices were calculated on each, which led to a similar solution with the three-cluster solution still found optimal by a majority of the indices.[30] Additionally, I created silhouette plots for the two, three, five, twelve and fifteen-cluster solutions (three- and two-cluster solutions are shown in Figure 6.3). The three-cluster solution has the highest average silhouette width (except for the fifteen-cluster solutions) and the lowest number of negative values — which represent cases that are usually considered to have been sorted into the wrong cluster. To further validate this finding, both two- and three cluster solutions were interpreted as potentially describing frames. However, only the three-cluster solution is reported here as the two-cluster solution did not seem to make theoretical sense as both frame candidates appeared highly heterogeneous and were hard to interpret.

Figure 6.4 shows the cluster means for the three-cluster solution. Each column contains the mean for all articles in one cluster. This means that high values in Figure 6.4

[30] Loosely following the robustness check suggested by Everitt (2011, pp. 269-271).

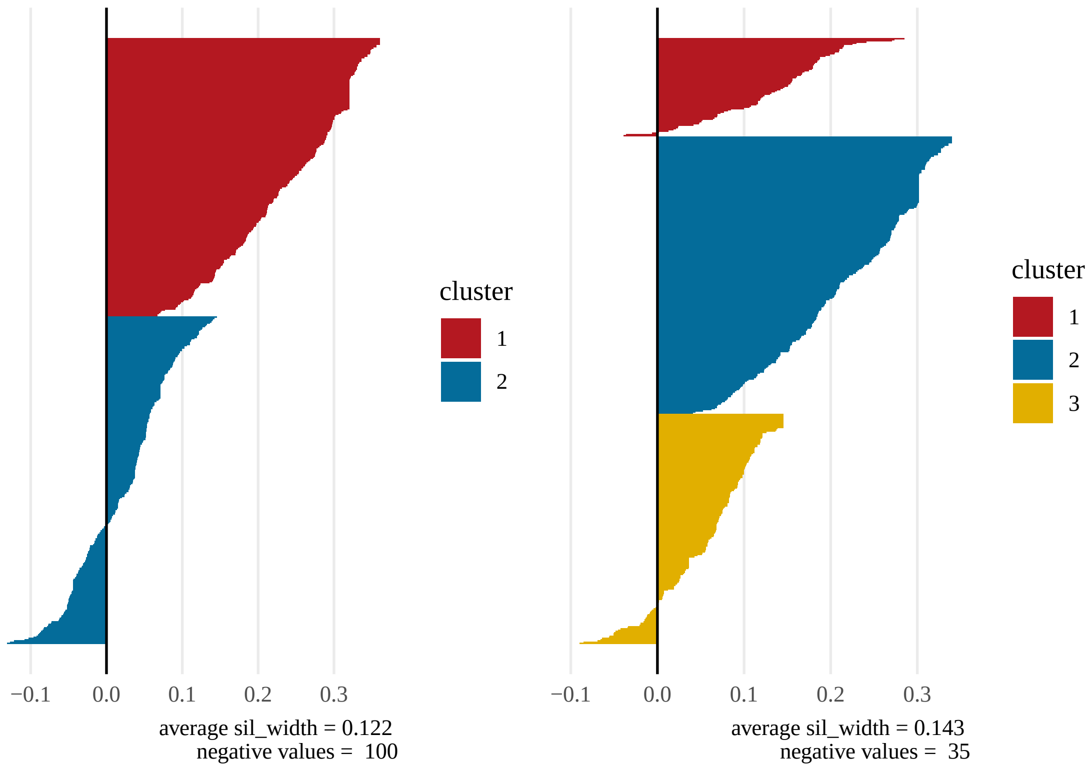


Figure 6.3: Silhouette Plots for 2- and 3-Cluster-Solution

indicate that a code was used more often, whereas low values indicate it was used less often globally.[31] The cluster means are meant to highlight the defining features of a frame. Again, frames are seen here as recurring pattern of the same frame elements used in a number of different articles. As cluster analysis groups together articles using the same frame, the mean of a cluster can describe which elements are used often, allowing to infer the features of the dominant frame in a cluster. Three things are,

[31] As the values for each code in each article are already aggregated from the values in the paragraphs, the cluster means are not directly interpretable. Only the two extreme values — 0 and 1 — can be directly explained. A cluster mean for a code would be 0 if none of the paragraphs in the articles in a cluster would use a code and 1 if all paragraphs in all articles in the cluster use the code. However, the means can still be interpreted as we know that higher mean values indicate that a code was assigned in articles in this cluster more often. For example: a hypothetical cluster contains only two articles. In one of them violence and event appear as the topic in 50% of paragraphs. In the second article, the distribution is 0.4 for Violence and 0.6 for Event. Overall, the cluster means for topic would be 0.45 for Violence and 0.55 for Event, while being 0 for all codes that do not appear in any articles in the cluster. Note, that the cluster means for topic, actor, risk and so forth do not necessarily add up to 1. As judgement, for example, is not often assigned, most articles will have a mean of 0 for both negative and positive judgement. A sum of less than 1 for the cluster means of a variable, therefore, indicates that some articles in the cluster did not contain any code for topic, actor, risk and so forth.

	Troublemakers (16.2%)	Struggle for (Just) Cause (45.8%)	Mixed (38.0%)
Problem Definition: Topic: Clash	0.36	0.043	0.122
Problem Definition: Topic: Violence/Crime	0.262	0.015	0.046
Problem Definition: Topic: Cause	0.047	0.305	0.094
Problem Definition: Topic: Protesters	0.034	0.06	0.071
Problem Definition: Topic: Spectacle	0.016	0.096	0.058
Problem Definition: Topic: Protest tactics*	0.019	0.014	0.017
Problem Definition: Topic: Policing tactics*	0.061	0.016	0.099
Problem Definition: Topic: Response*	0.005	0.013	0.014
Problem Definition: Topic: Confrontation/Showdown	0.039	0.018	0.109
Problem Definition: Topic: Public opinion	0.007	0.018	0.007
Problem Definition: Topic: Nuisance	0.008	0.021	0.017
Problem Definition: Topic: Other activism*	0.003	0.022	0.012
Problem Definition: Topic: Judicial prosecution*	0.047	0.024	0.117
Problem Definition: Topic: Effect of protest*	0.006	0.026	0.091
Problem Definition: Topic: Event	0.085	0.309	0.111
Problem Definition: Actor: Protesters	0.478	0.925	0.102
Problem Definition: Actor: Police	0.301	0.018	0.151
Problem Definition: Actor: Officials	0.012	0.016	0.105
Problem Definition: Actor: Business	0.007	0.014	0.027
Problem Definition: Actor: Other Political Elite*	0.01	0.011	0.033
Problem Definition: Actor: Other*	0.01	0.012	0.026
Causal Attribution: Benefit Attribution: Police	0.133	0.013	0.046
Causal Attribution: Benefit Attribution: Protesters	0.048	0.043	0.01
Causal Attribution: Risk Attribution: Protesters	0.386	0.095	0.129
Causal Attribution: Risk Attribution: Police	0.372	0.007	0.016
Causal Attribution: Risk Attribution: Officials*	0.027	0.156	0.057
Causal Attribution: Risk Attribution: Other*	0.058	0.015	0.03
Moral Evaluation: Benefit: Reinstating public order	0.193	0.015	0.041
Moral Evaluation: Benefit: (Just) Cause*	0.024	0.079	0.031
Moral Evaluation: Risk: Grievance*	0.018	0.27	0.049
Moral Evaluation: Risk: Public safety	0.835	0.018	0.016
Moral Evaluation: Risk: Property destruction	0.052	0.047	0.037
Moral Evaluation: Risk: Breaking laws*	0.011	0.028	0.081
Moral Evaluation: Risk: Decay of morals or other social norms	0.01	0.009	0.037
Moral Evaluation: Risk: Nuisance	0.015	0.064	0.054
Moral Evaluation: Risk: Harm discussion*	0.007	0.026	0.046
Moral Evaluation: Risk: Trivializing (political) discussion	0.02	0.031	0.041
Moral Evaluation: Risk: Bad for business*	0.002	0.006	0.034
Moral Evaluation: Risk: Costs of demonstrations	0.027	0.016	0.055
Moral Evaluation: Risk: Suppression/Censorship	0.003	0.023	0.045
Treatment: Judgement: Negative	0.048	0.042	0.103
Treatment: Judgement: Positive	0	0.043	0.031

Figure 6.4: Heatmap of Cluster-Means for all Codes

therefore, important to keep in mind when interpreting Figure 6.4: first, globally high values, which indicate that one code was used in a large proportion of the articles in one cluster and/or was usually not challenged by frequent occurrence of another code in the same frame element; second, comparatively high values, which indicate that while a code was not used predominantly, it was still used more often in one frame than in the others; third, low values, which mean a particular code was mostly absent from one cluster of articles. To make interpretation easier, Figure 6.4 is presented as a heatmap, where colours indicate globally high (red) and low (yellow) values, making it easier to find the most distinctive features of a cluster.

The three clusters were named *Troublemakers*, *Struggle for (Just) Cause* and *Mixed* after their most distinctive features. The ***Troublemakers*** cluster is characterised by the highest values in the two topics *clash* and *violence/crime* which together have a mean of 0.621. The vast majority of articles mention either protesters (0.478) or the police (0.301) as the main actor. The main *risk* identified by this cluster is *public safety* (0.835), while *property destruction* is also still relatively high compared to the other clusters (0.052). Interestingly, the risks are attributed not only to the protesters (0.386), but almost as often, the police takes the blame (0.372). In terms of *treatment*, the negative judgement is surprisingly low at (0.048), which is nearly the same as in the second and smaller than in the third cluster. There is, however, no positive judgement whatsoever. However, even without explicit judgement, the selection of codes in the variables clearly indicates that protests are usually seen negatively by articles in this cluster.

The frame *Troublemakers*, therefore, mirrors, to a certain extent, the description of the coverage in the *protest paradigm* literature. As suggested by the literature, there was little room for discussing the cause (coded as topic *topic: cause*) of a protest as articles in this cluster mostly focus on the negative aspects of a protest. Although the cause was not entirely absent (0.047), it appeared far less often than in the other clusters and the code *risk: grievance* was almost never mentioned.

The cluster dubbed ***Struggle for (Just) Cause*** is characterised by the highest value in *topic: cause*, which was assigned when articles describe why people were protesting and what they were protesting for/against. In other words, the coverage the *protest paradigm* literature suggests is missing from the discussion of protest. With a mean of 0.305, the reasons behind a protest are mentioned much more often in this cluster than in the other two. At the same time, the topic *event* has a value which is almost as high (0.309) and *topic: spectacle*, which describes eye-catching stunts or involvement of prominent figures, is still relatively common compared to the other clusters (0.096). Other topics are notably rare. Protesters are almost always portrayed as the actor in this cluster (0.925), apparently signalling a heightened degree of agency. Together, this suggests a cluster of articles which focuses on the actions *and* motivation of protesters.

The same focus is continued in the *moral evaluation* and *causal interpretation* codes: The main benefit is *(just) cause* (0.079) on a level 2-3 times higher than in the other clusters. By far the most prominent risk is *grievance* (0.27), which basically means that many articles discuss the main grievance identified by protesters. Officials are most often seen to be responsible for these risks (0.156) while protesters are made responsible for the other risks.

Most of the positive judgement that was assigned in the coding process ends up in this cluster, although it is still at an overall low level at just 0.043. Surprisingly though, the code negative judgement is as common (0.042). Overall then, most of the cluster means point to a framing that describes the cause of the protesters' actions, sometimes along with the struggle they face to eliminate their grievances.

The final cluster was dubbed ***Mixed***. Its most distinct feature is that it does not have any distinct features. Codes are more or less evenly split, especially in the *topic* category. None of the top topics, *clash* (0.122), *judicial prosecution* (0.117), *event* (0.111) and *confrontation/showdown* (0.109), have particularly high global means, while the cluster means for the remaining topics, such as *cause* (0.094), are almost at the same level. The main actor variable is also almost evenly split between *police* (0.151), *of-*

ficials (0.105) and *protesters* (0.102). Somewhat surprisingly, while none of the risks have globally high values, eight of the eleven risks have the highest mean value in this cluster, even though always by a narrow margin. The two risks which were defining for the other two clusters, *public safety* and *grievance*, are rather uncommon in the *Mixed* cluster though. The cluster also shows the highest negative judgement (0.103), which is still on a low level globally, however.

The cluster means for the *Mixed* cluster, therefore, suggest two different interpretations. One is that articles in this cluster attempt a balanced view of the situation, combining elements of the other two clusters, devoting a similar degree of attention to both sides, the cause of an event along with mentioning numerous risks. The second one is that this is a residual category, comprised of articles that do not fit anywhere else. One indication for the latter interpretation is that the topic *judicial prosecution*, which often occurs isolated from other common frame elements, has its highest mean in this cluster. If an article contains few codes, this makes it dissimilar to articles in either one of the other clusters which usually have several non-zero values. Looking back at Figure 6.3, it can be seen that this cluster (cluster 3) had the most cases with negative silhouette widths, indicating the poorest fit among the three clusters.

Summing up the **frame identification** results from the cluster analysis, it is possible to extract two coherent frames. These two frames intuitively make sense, despite minor inconsistencies. For the third frame, there are two competing, yet reasonable interpretations. Furthermore, going back to the literature, McLeod and Hertog (1999) describe strikingly similar broad frame categories: they differentiate between marginalizing coverage, which is similar to the *Troublemakers* cluster; sympathetic and balanced coverage, which can be found in the *Struggle for (Just) Cause* cluster; and mixed coverage, which fits the description of the *Mixed* cluster above.

Nevertheless, due to some of the described inconsistency, especially in the *Mixed* cluster, some doubt remains whether the approach is also suitable for the next analysis step, *frame coding*. Since cluster membership, and thus the prediction of the main framing

Table 6.6: Confusion Matrix. k-Means vs. Manual Coding of Main Frame in an Article (Agreement in Green; Disagreement in Red)

	manual		
k-means	**1**	**2**	**3**
1	32	1	5
2	23	68	29
3	44	11	38

metric	value
Accuracy	0.55
Average Precision	0.61
Average Recall	0.57
Average F1	0.54

by the k-means algorithm, is available at this point, it seemed convenient to validate the results and gain some further insights into the identified frames by comparing clustering against human judgement. To do so, half the coded sample (n = 250) was assessed manually again and each sorted into one of three classes. I used the description of frames presented in this section to make decisions about which frame, *Troublemakers*, *Struggle for (Just) Cause* or *Mixed*, is the most prominent in the coded articles.

Table 6.6 shows the agreement between the machine and human coder in a confusion matrix. Values highlighted in green show where human coding and classification by the machine came to the same assessment about the most prominent frame, values highlighted in red show where one disagrees with the other. The validation indicates that the level of agreement is not acceptable as human and computer coder disagree in almost half of the assessed cases. Specifically, k-means underestimates the number of articles in cluster 1 (*Troublemakers*). According to k-means, *Troublemakers* is the smallest cluster making up 16.2% of the coded articles, while *Struggle for (Just) Cause* and *Mixed* command bigger shares (45.8% and 38.0% respectively). In contrast, from the manual coding, *Troublemakers* emerges as the biggest cluster with 39.4% of coded articles falling in this category, followed by *Struggle for (Just) Cause* (31.9%) and *Mixed* (28.7%) which is the smallest now. It is also noteworthy that the manual coding showed that the *Mixed* cluster can indeed be split into two categories that make theoretically more sense: one frame that can be described as balanced and a second residual one, which has few and isolated codes, marking articles in which protest is described rather casually.

This does not mean that k-means delivers "wrong" results. The algorithm does fulfil its purpose to statistically group articles which are most similar according to the selected variables — that is, the codes. And as the next section will show, the cluster means for *Troublemakers* and *Struggle for (Just) Cause* do describe overarching framing patterns. However, the issue remains that classification of individual articles does not coincide with human judgement. Crucially, *frame coding* has not been reported in previous research which used cluster analysis of frame elements (David et al., 2011; Matthes and Kohring, 2008), meaning that guidance on how to proceed with difficult results of the clustering step is unavailable.

This thesis is thus a first step towards guiding future research on how to use dimension reduction techniques for framing analysis. The contribution of this section was thus to show that even a carefully executed frame identification procedure based on cluster analysis was not able to also perform frame coding, even though the identified frames appear meaningful. Researchers using this technique should, therefore, be sceptical about the proportion of measured frames in a corpus. In the next section I will introduce factor analysis as an alternative dimension reduction technique that is better suited for frame identification and frame coding both theoretically and in terms of performance. The strengths of factor analysis are thus established in direct comparison to the cluster analysis results.

6.2.3 Factor Analysis

The principle of factor analysis is that variables which correlate — in this case codes which are often used together — are determined by underlying latent dimensions — in this case the framing of an article. The goal of the technique is thus to uncover underlying factors in order to interpret them better, or to work with the reduced number of dimensions directly. Whereas clustering attempts to find the best grouping for cases, in contrast, factor analysis tries to group variables. Factor analysis is, therefore, a suitable alternative or complementary technique for clustering here, with the ultimate goal to get a clearer understanding of the emerging latent dimensions. These factors can then be used to identify frames in articles using factor score estimates, which represent a case's placement on the factors. A crucial difference between factor and cluster analysis is that the latter allows mixed membership of cases, meaning that each article may contain several frames. In the following, I describe the outcome first of an exploratory factor analysis (EFA) — a purely data driven technique — and a confirmatory factor analysis (CFA) — which tests theoretical assumptions about underlying factors. I use insights from EFA, knowledge about the codes, as well as the outcome of the clustering, to test a theory about the real underlying factors through CFA — after which the tested factors are regarded frames.

As EFA assesses the correlation between variables, only codes which correlate with other codes should be employed in the analysis. I, therefore, initially remove codes which do not correlate with any other code (i.e., only codes remain with one r either $r > 0.3$ or $r < -0.3$) before analysis. Out of the 42 codes, this leaves 25 variables suitable for the EFA. Most of the excluded codes appear in the data rarely. However, this criteria also excludes some important codes, notably *topic: spectacle* and *risk: property destruction*.[32] This means that the EFA model does not depict the entire picture for now. For that reason, I progress in two steps: I first compute a factor analysis model for a selected set of statistically suitable variables before repeating the analysis on the entire set of variables. By comparing the results, I can make sure that the retrieved factors are robust while at the same time including all coded material.

Conducting a standard collection of feasibility tests for the reduced model, I found that the subset of the data is suitable for EFA: Bartlett's test succeeded with $p < 0.001$. The overall Kaiser-Meyer-Olkin (KMO) measure for the dataset is 0.58, which translates to mediocre sampling adequacy. The determinant for the dataset is 0.0022, which is considered a good value as 1 would represent completely unrelated variables

[32] Taking a closer look at those codes revealed that the reason why they do not correlate with other codes is that they are used throughout very different articles that would appear in all of the above described clusters indiscriminately.

and 0 would indicate a singular matrix. Good values are small but above 0.00001, which is the case (Field et al., 2012, pp. 769-772).

As with cluster analysis, finding an optimal number of factors is the first step in exploratory factor analysis. Figure 6.5 shows a scree plot, which displays the number of factors on the x-axis and their respective eigenvalues on the y-axis. Additionally Figure 6.5 shows how much of the variance in the data is explained by each factor. The number of factors is determined at the point, beyond which the remaining eigenvalues and the variance explained are all relatively small and of comparable size (Jolliffe, 2002, pp. 115-118). In other words, we look for a bend or “elbow” where the curve gets abruptly flatter. The clearest elbow in Figure 6.5 is at 10 factors, and as the rule is to remove all eigenvalues after and including the elbow, the number of factors for the analysis was determined to be nine (Jolliffe, 2002, pp. 115-134). Using nine factors, the model explains just above 64% of total variance in the data, which is not great but is plausible for the purpose here.

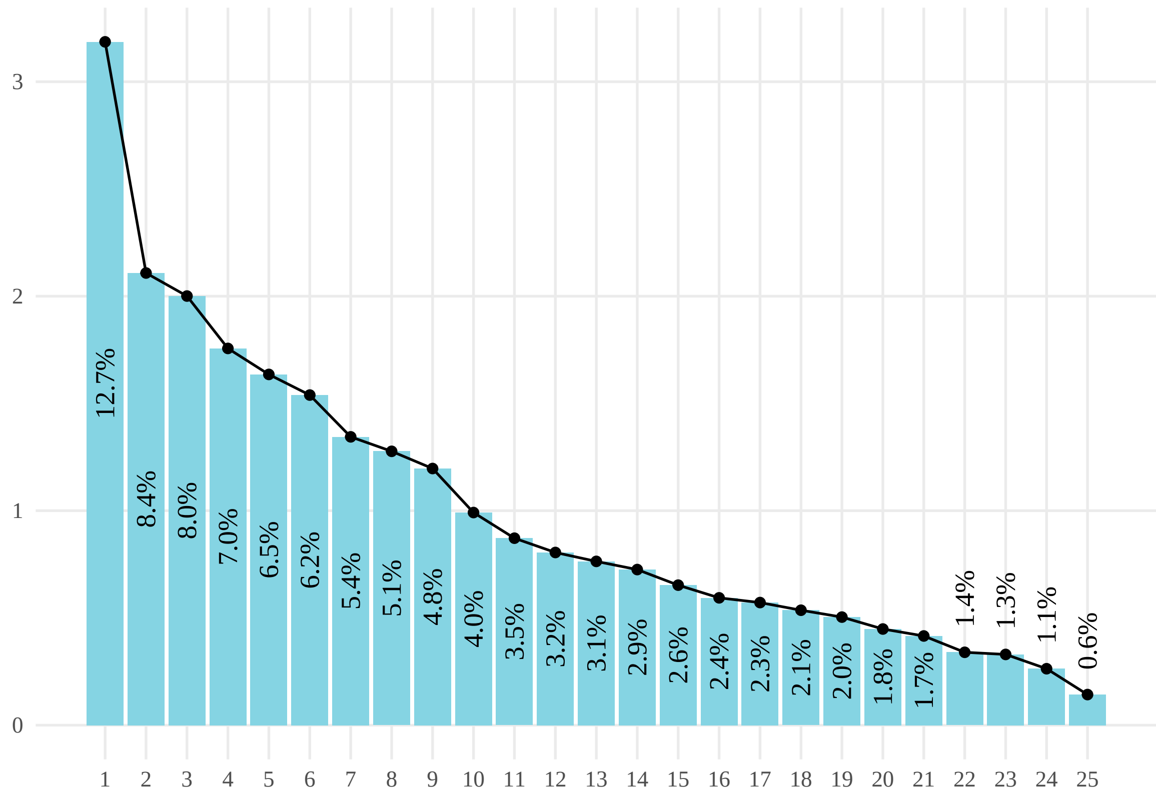


Figure 6.5: Eigenvalues for Number of Factors and Percentage of Explained Variance

In addition, a *parallel* analysis was performed using the R package *psych* (Revelle, 2019), which confirmed the choice for nine factors as the best solution. This technique, which is a more automatic approach to determine an optimal number of underlying dimensions, compares the scree of factors of the observed data with that of a random data matrix of the same size as the original. In doing so, the eigenvalues of the real data are compared with eigenvalues from a data set without underlying factors (Horn, 1965).

EFA was performed using again the package psych (Revelle, 2019) with oblique rotation, which is suggested if factors cannot be assumed to be uncorrelated.[33] Table 6.7 shows the factor loadings for the model as well as model fit measures in the table footnote. The overall fit of 0.84 and root means squared residual of 0.07 are not great but still suitable for the purpose of developing an empirically sound category system of frames. Less than half the variables have residuals larger than 0.05, which is considered to indicate good fit.

To interpret the factors, we look at the factor loadings in Table 6.7. High values indicate that a code is important for a given factor. As factor loadings below 0.3 are usually not interpreted, they are omitted here in favour of better readability. The factors are ordered left to right from most to least important in terms of how much of the total variance in the data they explain. Factors were then named based on the knowledge about the codes with high loadings in each factor. We see that the first and most important five of the nine factors link up with either the *Troublemakers* or *Struggle for (Just) Cause* cluster identified in the last section. This is reassuring of the basic validity of the cluster means as discussed above. However, the EFA offers greater detail and we can see responsibility and blame for trouble and causes are split between actors. One potential problem of the *Troublemakers* cluster was that it does

[33] This is the case here as many of the categories are mutually exclusive in the sense that a high value on, for example, the *topic: clash* means that there can be only a low value in any other topic. A factor with a high loading in *topic: clash* is, therefore, likely negatively correlated with factors that include a high loading in another topic.

Table 6.7: Factor Loadings

	1. Trouble: police good	2. Trouble: police bad	3. Cause: officials bad	4. Cause: protesters good	5. Trouble: protesters bad	6. Bad for business	7. Business response	8. Trivial discus-sion	9. Nuisance
Problem Definition: Topic: Clash	0.792								
Causal Attribution: Risk_Attribution: Police	0.687								
Problem Definition: Actor: Police	0.591	0.337							
Moral Evaluation: Risk: Public safety	0.442				0.601				
Causal Attribution: Benefit_Attribution: Police		0.941							
Moral Evaluation: Benefit: Reinstating public order		0.935							
Moral Evaluation: Risk: Grievance			0.77						
Problem Definition: Topic: Cause			0.735						
Causal Attribution: Risk_Attribution: Officials			0.7						
Moral Evaluation: Benefit: (Just) Cause				0.886					
Causal Attribution: Benefit_Attribution: Protesters				0.762					
Treatment: Judgement_Positive: 1				0.613					
Problem Definition: Topic: Violence/Crime					0.88				
Causal Attribution: Risk_Attribution: Protesters					0.611			0.324	
Problem Definition: Topic: Effect of protest						0.757			
Moral Evaluation: Risk: Bad for business						0.693			
Problem Definition: Actor: Business							0.839		
Problem Definition: Topic: Response							0.817		
Moral Evaluation: Risk: Trivializing (political) discussion							0.388	0.651	
Problem Definition: Topic: Protesters								0.833	
Treatment: Judgement_Positive: 0								0.36	0.358
Problem Definition: Topic: Nuisance									0.798
Moral Evaluation: Risk: Nuisance									0.758
Problem Definition: Actor: Protesters									
Problem Definition: Topic: Event									

* Root mean squared residual = 0.07; proportion of absolute residuals > 0.05 = 0.43; overall fit = 0.84

not differentiate if protesters or police were blamed for a clash, whereas the EFA paints a more differentiated picture.

The remaining four factors combine codes which portray protest as a nuisance in varying degrees: two factors focus on the downside for businesses, either by directly being *bad for business* or by *trivializing the discussion*, a code which was usually assigned if protesters were accused of misunderstanding the reasons behind the decision or policy they oppose. This code was also responsible for the name of the next factor, *Trivial Discussion*, which has the highest loading for this code and combines it with a negative judgement and the topic *protesters*, which was assigned when articles focus on the appearance, mental ability, visual deviance and oddities of the protesters. The last, and least important, factor has the highest loading in the topic and risk nuisance, which were assigned when a text focuses on the inconvenience a protest caused for citizens or the government. Two codes at the bottom of Table 6.7 did not have any factor loading above 0.3, which means they are apparently not important in explaining variance in the data.

In sum, the results of the EFA provide more insight into the meaning of the clusters identified in the last section. Generally, it appears that while clustering did identify frequently recurring patterns, factor analysis provides a clearer image on what those patterns are really about in terms of framing. The outcome of the analysis using the two techniques thus differs in two ways: Firstly, the statistically determined optimal number of factors is larger than the number of clusters, which means the results show a greater level of detail — although several of the factors seem rather similar and in some cases extremely specific. Secondly, there is no mixed category, which is due to the design of EFA, a mixed membership model, which allows that cases that include codes from multiple dimensions can have membership to multiple factors at once.

As mentioned above, the EFA model in Table 6.7 only makes use of the 25 variables which satisfy the statistical requirements to ensure a valid factor analysis. However, as described above, this means that a relatively large portion of the data generated through coding is not taken into account. A second model was thus set up to include all coded variables. Like for the reduced model, several standard feasibility tests were performed: Bartlett's test still succeeded with $p < 0.001$ and the determinant is still small and above 0.00001, yet barely this time. The KMO measure on the other hand is problematic in the full model with a value of just 0.12 which represents a warning that factor analysis might be inappropriate (Field et al., 2012, pp. 769-772). This warning is ignored for now though as the model is validated against the reduced model above.

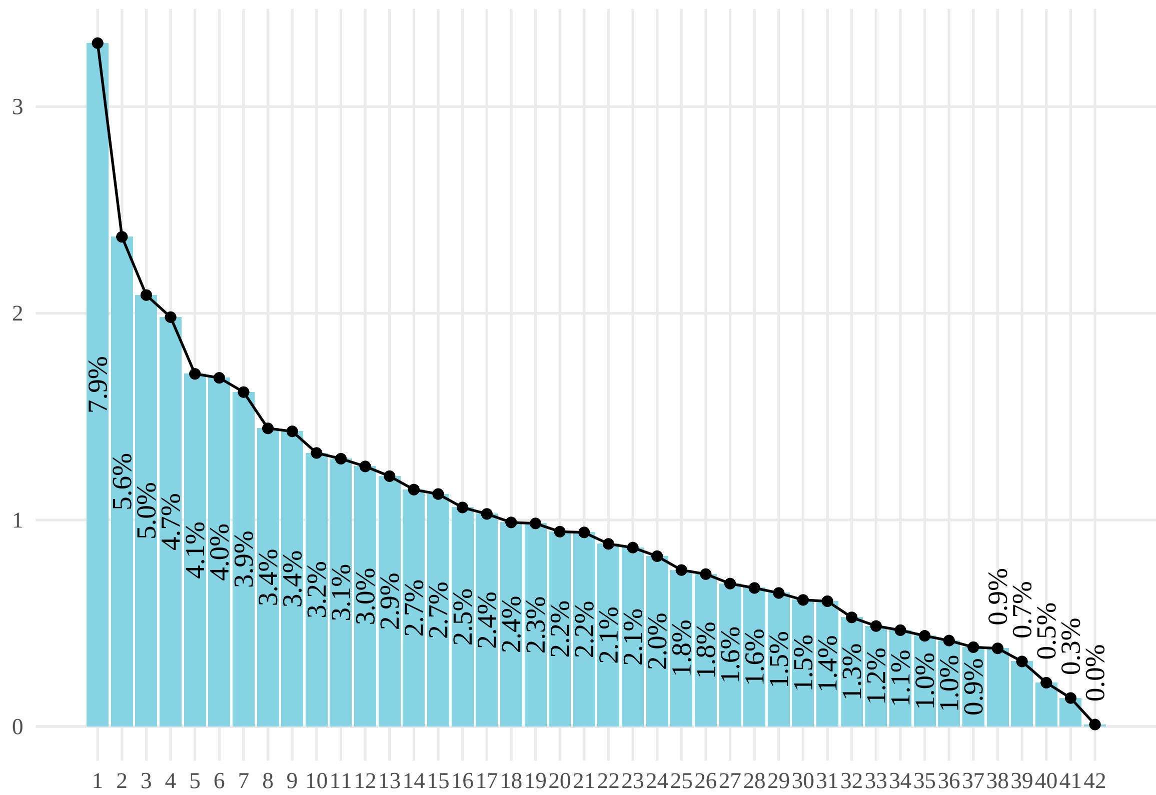


Figure 6.6: Eigenvalues for Number of Factors and Percentage of Explained Variance for the Full Model

Figure 6.6 shows the same plot as above but with the full dataset. Again, nine factors appear as a good choice as there is a substantial drop to ten before the curve flattens. Nine factors explain a total of 42% of total variance. Again a *parallel* analysis was performed as well using the R package *psych* (Revelle, 2019) which suggested 13 factors instead, explaining total variance of 54%. As the *parallel* analysis is usually considered

Table 6.8: Factor Loadings Full Model

	1. Trouble: police bad	2. Trouble: police good	3. Cause: officials bad	4. Cause: protesters good	5. Trouble: protesters bad	6. Decay of morals	7. Business response	8. Bad business	9. Idiots at large	10. Policing tactics	11. Harassment	12. Nuisance	13. Other actors
Causal Attribution: Risk_Attribution: Police	0.751												
Problem Definition: Topic: Clash	0.699												
Moral Evaluation: Risk: Public safety	0.656				0.37								
Problem Definition: Actor: Police	0.449	0.32											
Causal Attribution: Benefit_Attribution: Police		0.938											
Moral Evaluation: Benefit: Reinstating public order		0.933											
Moral Evaluation: Risk: Grievance			0.763										
Problem Definition: Topic: Cause			0.733										
Causal Attribution: Risk_Attribution: Officials			0.704										
Moral Evaluation: Benefit: (Just) Cause				0.865									
Causal Attribution: Benefit_Attribution: Protesters				0.727									
Treatment: Judgement_Positive: 1				0.64									
Problem Definition: Topic: Violence/Crime					0.725								
Causal Attribution: Risk_Attribution: Protesters					0.629								
Moral Evaluation: Risk: Property destruction					0.612								
Treatment: Judgement_Positive: 0					0.336								
Problem Definition: Topic: Confrontation/Showdown						0.734							
Moral Evaluation: Risk: Decay of morals or other social norms						0.674							
Problem Definition: Actor: Other Political Elite						0.629							
Problem Definition: Actor: Officials						0.311				0.324			
Problem Definition: Actor: Business							0.821						
Problem Definition: Topic: Response							0.773						
Moral Evaluation: Risk: Trivializing (political) discussion							0.373		0.663				
Problem Definition: Topic: Effect of protest								0.809					
Moral Evaluation: Risk: Bad for business								0.698					
Moral Evaluation: Risk: Costs of demonstrations								0.399					
Problem Definition: Topic: Protesters									0.83				
Problem Definition: Topic: Policing tactics										0.795			
Moral Evaluation: Risk: Suppression/Censorship										0.557			0.365
Problem Definition: Topic: Event											0.553		
Problem Definition: Actor: Protesters											0.454		
Moral Evaluation: Risk: Harm discussion											0.338		
Problem Definition: Topic: judicial prosecution											-0.662		
Moral Evaluation: Risk: Breaking laws											-0.413		-0.367
Problem Definition: Topic: Nuisance												0.77	
Moral Evaluation: Risk: Nuisance												0.748	
Causal Attribution: Risk_Attribution: Other													0.583
Problem Definition: Actor: Other													0.545
Problem Definition: Topic: Spectacle													-0.403
Problem Definition: Topic: Other activism													
Problem Definition: Topic: Protest tactics													
Problem Definition: Topic: Public opinion													

[*] Root mean squared residual = 0.06; proportion of absolute residuals > 0.05 = 0.4; overall fit = 0.74

more advanced and reliable, the model was calculated with 13 factors, again using the package psych (Revelle, 2019) and oblique rotation.

Results from the EFA with the full dataset and 13 factors are shown in Table 6.8. As becomes apparent, the results are actually very similar to the model with the reduced dataset. The first five factors, as well as number seven, are actually exactly the same. A few factors have changed places: factor 6 (*Bad for business*) is now factor 8; factor 8 (*Trivial discussion*) is 9 and factor 9 (*Nuisance*) is 12 now. The factors 6, 10, 11 and 13 are new.

Note that factors are ordered again left to right from most to least important in terms of how much of the total variance in the data they explain. The fact that the five most important factors are the same as before suggests that this model arrives at largely the same conclusion about important latent dimensions in the data. The most important five are also, as the names suggest, closest to what we have seen from the cluster analysis, albeit they highlight responsibility for the problems highlighted in these frames. This is encouraging as it indicates that we see real latent dimensions here that have been uncovered using different approaches and even different data, if the reduced dataset which was used in the first EFA model is taken into account. Overall, the high degree of similarity between the model which was based on the reduced dataset and the model which makes use of the entire data is seen as evidence for the validity of the model that includes all data despite the relatively poor fit scores.

Additional to the first five factors, the remaining factors are more specific and, as mentioned above, might actually be too specific to be treated as frames. Therefore, the decisions was made to omit most of them. Only factor 1–6 and 12 appear fleshed out and meaningful enough from a theoretical view to qualify as frames. Additionally, a parsimonious set of categories is easier to interpret and ultimately *code* on the remaining data in the next section.

Yet omitting categories might undermine the validity of the model. This is where confirmatory factor analysis (CFA) comes into play. As mentioned above, EFA attempts to extract underlying dimensions from the data inductively. In contrast, CFA *tests* pre-specified relationships between observed measures and latent dimensions or factors. The model, which is a formal description of the expected relationships, is tested against the data and multiple competing models can be compared to determine which one fits best. This makes CFA suitable for the goal here, which is to make sure that omitting some of the extracted dimensions in order to arrive at a more parsimonious category system for the articles does not invalidate the model.

To perform the CFA I use the *lavaan* package in R (Rosseel, 2012). As the assumption of multivariate normality was not met, I used a diagonally weighted least squares (DWLS) estimator (Forero et al., 2009). Using CFA, I tested a model where only factors 1–6 and 12 remained. The comparative fit index (CFI) for this model is 0.94 which is above the 0.9 threshold usually deemed good (Brown, 2015, pp. 73-75). The adjusted goodness of fit index (AGFI) should be > 0.8 which is the case (0.87). The Root Mean Square Error Of Approximation (RMSA) is also good at 0.03 — RMSA values are considered good when smaller than 0.05, moderate if 0.05-0.1, and inadequate when larger than 0.1. The p-value threshold for close fit should be larger than 0.05, which is the case. The decision to include the factor *nuisance* was also comparatively confirmed to increase model fit.

It is, therefore, confirmed that the model is valid, even after removing some of the factors suggested by EFA. The remaining factors, 1–6 and 12, are thus treated as frames from here on. At this stage, the patterns of frame elements which are repeatedly used together have been rigorously tested: first broadly through the cluster means in the last section, and second in more nuanced fashion in the exploratory and confirmatory factor analysis. Consistently, categories highlight the violence attributed to protesters, the cause and issues represented through protest or the nuisance caused by protest. Through the triangulation of methods in this section, the identified frames are, therefore, regarded as plausible. The more fine-grained factors offer more detail and are more consistent, though, and were thus chosen for further analysis.

6.2.4 Interpretation of Frames

The short answer to RQ1 (*how do British newspapers frame the coverage of domestic protest events?*), based on the analysis explained above, is that there are seven main frames used by mainstream news media to cover domestic protest events. To recap, the basis of this conclusion are 500 randomly chosen articles from the dataset containing items published in the 8 selected major newspapers, spanning the time frame 1992-2017.

Specifically, these seven factors were named *trouble (police bad)*, *trouble (police good)*, *cause (officials bad)*, *cause (protesters good)*, *trouble (protesters bad)*, *decay of morals*, and *nuisance*. The naming convention was chosen to highlight how close results from the cluster and factor analysis turned out to be. However, these names are also rather unwieldy for further discussion of analysis results. To make it easier for readers to follow the analysis, frames are given new names at this point to highlight only their most prominent features. Figure 6.7 serves as a visual aid for the discussion but features the same values as Table 6.8.

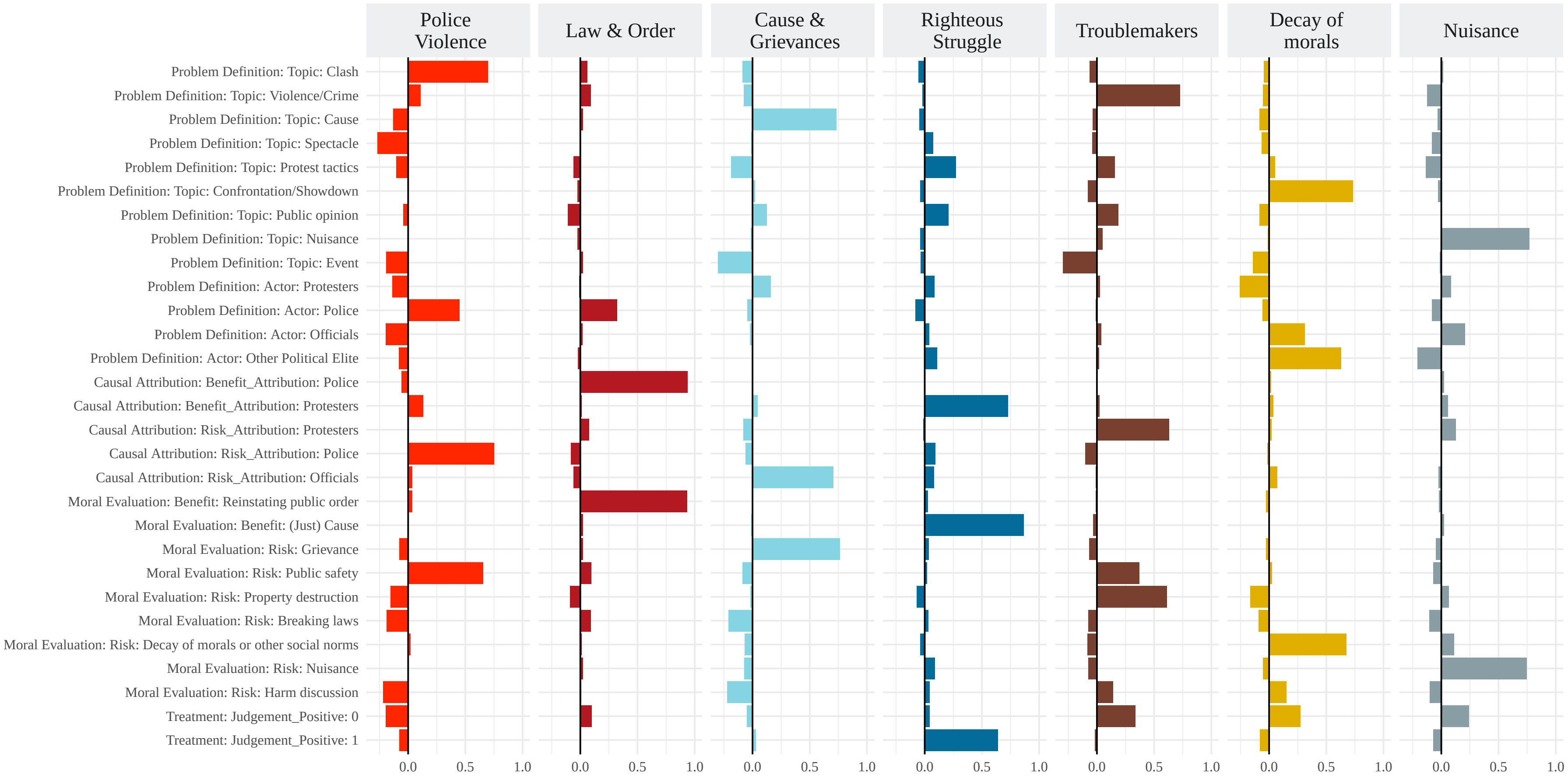


Figure 6.7: Factor Loadings for Selected Frames

In this section, I will describe in detail each of the frames and how they relate to the literature. The first frame is ***police violence*** (previously called *trouble* (*police bad*)). This frame highlights the clash between protesters and police and the risk this poses to members of the police force, the protesters and the general public. In some ways, this matches the existing literature. According to a range of studies, the emphasis on violence is a key component of the *protest paradigm*, which consistently found that clashes with the police and other violence is one of the main elements the media highlight, even when little to no violence actually occurs (e.g., Dardis, 2006a; Gitlin, 1980; Kilgo et al., 2019; Lee, 2014; McFarlane and Hay, 2003; McLeod, 1995, 2000; McLeod and Hertog, 1999; Mourão and Chen, 2019; Murdock, 1973; Shahin et al., 2016; Smith et al., 2001; Weaver and Scacco, 2012). Contrary to most theoretical expectations, however, the *police violence* frame portrays police, who are most often described as the main actor, as responsible for the trouble — hence its name. It is thus questionable if this particular frame causes the delegitimising, marginalising, and demonising effect commonly suspected by previous studies. On the one hand, the main negative effect — diverting attention from the cause of a protest event to its form — still takes place. On the other hand, the frame *police violence* can legitimise protesters struggle and help to shape public opinion about a protest group to their advantage. This has been found to be the case before (Cammaerts, 2013), and is possibly even more prevalent when journalists are the victims of police violence (Araiza et al., 2016).

One prominent example in which this frame was frequently employed in the coded sample was when Ian Tomlinson died from wounds inflicted by a police officer during the 2009 G20 summit protests in London. In the aftermath of Tomlinson's death, a broader discussion about the heavy-handedness of police during the protests and protests in general arose, often quoting individual protesters and bringing some protest groups to prominence. While this might not have aided the G20 protests in particular, unprovoked police violence can generate attention for a group beyond the individual event, as it happened for black-led protests in the 1960s (Wasow, 2020), the indignados

(Gerbaudo, 2012), during the Egyptian revolution (Harlow and Johnson, 2011) and, at least to some degree, during the Gezi park protests in Turkey (Oz, 2016). On the other hand, the frame was also sometimes used for May Day and student protest, when the police was criticised for 'loosing control' of the situation by not using more extreme tactics. This, of course, only complicates the interpretation of this frame further.

The second frame, ***law & order*** (previously called *trouble* (*police good*)), is the theoretical opposite of the first. It highlights the positive actions of the police, which are usually deemed to reinstate public order during a "rowdy" protest. It is, therefore, closely related to *police violence*, even though the respective factor does not directly reference the clash between protesters and police. To demonstrate this in an example:

> "But Tory MP Peter Bottomley said: 'Officers were fully justified in making sure the crushing stopped. I believe they saved many serious casualties, if not fatalities' " (Mackay and Yates, 2004-09-16).

This frame is more in line with previous findings and, when employed, it contributes to legitimising police action and even violence against protesters (e.g., Brasted, 2005; Kilgo and Harlow, 2019; McFarlane and Hay, 2003; McLeod and Hertog, 1992, 1999).

The delegitimising elements of reporting are most prominent in the third frame, ***troublemakers*** (previously called *trouble* (*protesters bad*)). Reports employing this frame highlight the violence caused by protesters and the associated risks, especially property destruction and risks for public safety. Protesters are deemed responsible for problems and the protest itself is often judged negatively. To give an example:

> "[Mrs May] defended the police, saying: 'I want to be absolutely clear: the blame for the violence lies squarely and solely with those who carried it out. The idea, that some have advanced, that police tactics were to blame when people came armed with sticks, flares, fireworks, stones and snooker balls, is as ridiculous as it is unfair.' " (Fresco and Ford, 2010-12-14).

This frame is one of the most prominently described in the literature on protest coverage. Specifically, this frame has been labelled "riot" or "violence" frame and is featured in almost all studies about protest coverage (e.g., Chan and Lee, 1984; Gitlin, 1980; Kilgo and Harlow, 2019; McFarlane and Hay, 2003; McLeod and Hertog, 1999; Murdock, 1973). For the UK, previous studies have found this to be a common frame, even when violence is only a minor part of an actual event (Dardis, 2006b; Gavin, 2007, 2010; Murdock, 1988).

The concern motivating many studies about news media coverage of protests is that the media might divert attention away from the substantive content and background of a protest to its form (Coombs et al., 2020; Kilgo and Harlow, 2019; McFarlane and Hay, 2003; McLeod and Hertog, 1999). In contrast, the next two frames highlight protesters' messages instead. ***Cause & grievances*** (previously called *cause* (*officials bad*)) does so by explicitly portraying the cause of a protest event as the main topic. The grievances of protesters are given a stage and officials are blamed for causing problems or not addressing them. Again, an example helps to illustrate this:

> "Why is this government so intent on killing off every institution we have? Small post offices are a lifeline to the elderly who cannot walk the distance to a main post office, and who often don't have cars, or cannot drive because of failing eyesight." (Emerson, 2008-03-04).

Some recent studies have shown that this type of coverage is not as uncommon as previously thought (Kilgo et al., 2019; Wouters, 2015b).

A more sympathetic coverage is also prevalent in the ***righteous struggle*** frame (previously called *cause* (*protesters good*)). The frame highlights the accomplishments of the protesters for a cause that is mostly deemed just and protesters receive positive judgement. To give an example:

> "Terry Deakin [...] said the problems had already cost him up to pounds 5,000 in lost business, but he still maintained the protesters had done a

> 'wonderful job. We have been in support of the drivers,' he said. 'It was short, sharp and got the message home' " (Griffin et al., 2000-09-16).

Yet, the articles using this frame provide less background and information about grievances. Instead, they sometimes highlight that the protesters represent what the majority thinks or highlight how peaceful and creative tactics have led to a positive image of a mentioned group. While the frame can thus legitimise a protest group, it does not necessarily provide space to air their grievances.

The frame ***decay of morals***, on the other hand, must be seen as another variation of the patterns described in the *protest paradigm* literature. The frame highlights the *topic: confrontation/showdown* and *risk: decay of morals or other social norms*. It thus falls in line with the notion that the arguments and demands of protesters are neglected in favour of routinised journalistic narratives. Confrontation/showdown means that a protest is just one event in an ongoing showdown between groups. Hence, the focus lies on an aspect of a protest that is thought to be most newsworthy: *conflict and controversy* — similarly to horse race rhetoric in reporting on political campaigning where every move by a politician is seen as just another step in a personal competition. Additionally, the frame captures a pattern of portraying protesters as a fringe group of freaks who disturb the social norms by questioning legitimate authority and disturb the general political consensus. This is best described in an example:

> "London and Glasgow will echo to the slogans of the morally deluded and the self-consciously caring. Papoose-wearers, manic recyclers, the priggish, the cranky, nudists and Woodcraft Folk will march this Saturday in a cloud of outrage. Peaceniks, marshalled by the Stop the War coalition, claim to march for the majority. They do not" (Millen, 2003-02-12).

The code from which the name of the frame was derived, *risk: decay of morals or other social norms*, originated in the study by Di Cicco (2010). He describes the moral aspect as follows: “obedience to legitimate authority is moral behavior and disobedience is both immoral and threatening to the social order” (p. 136). In other words, while protesters do not harm anyone physically, their actions are nevertheless harming society.

The last frame, ***nuisance***, played only a minor role in the results of the factor analysis. By highlighting the topic and risk *nuisance* this frame is linked to the kind of reporting described by Di Cicco (2010): it suggests that protest is a bothersome interruption of everyday activities, for example, through causing street closures, traffic jams or noise. Again giving an example:

> “St. Paul’s Cathedral will remain closed for the foreseeable future because of the anti-capitalist protest camp on its doorstep. Activists refused to budge yesterday — with many pledging to remain until Christmas and beyond. It came as thousands of Sunday worshippers were turned away. The London landmark shut its doors on Friday for the first time since the Second World War” (Sun, 2011-10-24).

This dismisses protest as a legitimate method of voicing a political opinion: the day to day activities of bystanders are deemed more important than the issues protesters oppose.

Considering the features of the described frames, it becomes clear that four of the seven factors that are considered frames have close links to the *protest paradigm*. Specifically, I consider *law & order*, *troublemakers*, *decay of morals*, and *nuisance* to be variations of the themes mentioned in the *protest paradigm* literature (McLeod and Hertog, 1999). The frames *cause & grievances* and *righteous struggle*, however, should be considered different from the reporting following the paradigm since their usage can be expected to legitimise protest or at least inform the public about specific the goals and background of events. The frame *police violence* cannot be described as either clearly legitimising nor delegitimising. Since the police is blamed for violence, protesters will usually appear as the "good guys". However, it is not clear if this necessarily helps protesters in achieving their goals, since it still diverts attention from the substantive content and background of a protest. In the further discussion, it is, therefore, treated as being between those two categories of frames.

6.3 Frame Coding

After frames were determined through dimension reduction techniques (*frame identification*), the final step was to code which frame or frames were used in each article in the full sample.[34] *Frame coding* was approached in three steps, which are described in the first part of this section: firstly, the factor analysis model was used to assess for each article in the manually coded sample if one or several of the frames was employed in the manually coded newspaper articles. Secondly, using this sample, machine learning models were trained and validated (Step 6 in Figure 4.1). Finally, the trained models were used to code the remaining set of newspaper articles about protest (Step 7 in Figure 4.1). The second part of this section presents the results from the *frame coding* step.

6.3.1 Teaching the Computer to Code Frames

To *code* frames in the database of newspaper articles ($N = 27,496$) supervised machine learning (SML) was used on the text data. To recap, the basic idea of SML is that a computer analyses the usage of words in a hand coded subset of the data, called *training set*, in order to infer or "learn" the implicit rules that led human coders to code a document in the way they did. These rules, referred to as a model, are then used to replicate the human coding on the remaining set of documents, called *test set* (Grimmer and Stewart, 2013).

[34] Note that the analysis this far had been done manually in a sub-sample of 500 randomly chosen articles.

Table 6.9: Example Factor Scores

Article ID	Police Violence	Law & Order	Cause & Grievances	Righteous Struggle	Trouble-makers	Decay of Morals	Nuisance
1323742	-0.327	-0.287	2.939	1.584	-0.382	-0.156	0.288
1302245	-0.326	-0.266	3.098	-0.258	-0.344	-0.241	-0.167
670056	-0.064	0.996	2.344	5.236	0.247	-0.239	-0.416
657629	0.857	-0.002	-0.418	-0.270	3.395	-0.287	-0.469
1181330	1.081	4.966	-0.432	-0.338	3.428	0.486	-0.027

For the analysis here, the *training set* consists of the same 500 randomly selected newspaper articles used to *identify* frames. The classes are derived from the factor scores which are available from the same factor analysis used to *identify* frames in the last section. Factor scores provide information about a case's placement on the factors. Higher scores mean a stronger affiliation of a case with a factor and hence a stronger presence of the frame in a text. As factor analysis employs a mixed membership approach, it can appreciate the existence of multiple frames in the same article, which is another advantage over a cluster analysis approach.

Table 6.9 shows factor scores for six articles and the seven selected factors as an example. From these scores, we can determine that there are three frames present in the first article (ID = 1323742): *cause & grievances*, *righteous struggle* and *nuisance*. The degrees to which these frames are present ranges from strongly (*cause & grievances*) to moderately (*nuisance*). Instead of using a new manually coded sample as training for SML, these scores can be used to determine the classes for texts, that is, the presence or absence of a frame in an article.

However, the currently available SML algorithms are designed to handle a specific task: a training dataset containing variables on cases and their respective class membership is fed to the algorithm for it to learn the relationship between the variables, in this case words, and it's class membership. The resulting model can then predict a probability for class membership of each case (Welbers et al., 2017).

This means that to use the factor analysis results as input classes, the standard approach had to be modified in two ways. First, to determine if a frame is present in an article I dichotomise the factor scores using a simple decision rule: a positive factor score represents the presence of a frame in an article, a negative score represents its absence. Secondly, as said, factor analysis employs a mixed membership approach, which allows that more than one frame can be present in an article. This makes theoretically more sense than assuming that each article employs exactly one frame and previous studies also assume that frames are not mutually exclusive (e.g., Cammaerts et al., 2013; Coombs et al., 2020). However, while most algorithms are able to handle classification of multiple categories, these need to be mutually exclusive. Due to the fact that each article can contain multiple frames, the approach had to be changed, using a common modification: instead of training one classifier to assess the membership of a case to a class, I train one model for each frame. This means that instead of one multi-class model, the task is split into seven binary classification models. Each of the seven models is then trained to predict if articles contain the respective frame or not. This is the suggested solution to mixed membership classification in computer science (Manning et al., 2008, pp. 281-283) and has also been used by Burscher et al. (2014) in a similar study.

In order to do statistical analysis on text data, words have to be turned into numbers first. This was done using the R software package *quanteda* (Benoit et al., 2018). I assume that documents are a *bag of words*, meaning the order of words is ignored for the analysis. Studies have shown that while intuition tells us that the order of words is important, and while it is simple to construct sentences in which this is true, these cases are rare in reality and do not usually hamper classification accuracy (Grimmer and Stewart, 2013).

Table 6.10 shows a small example of the data that is used for training. The rows in the matrix show each document, while the columns show each word or, more precisely, each feature. The values in the matrix show how often each feature was used in each document, except the values in the first and last column shown here. The information in the feature columns is used to train the model in order to predict the value of the

Table 6.10: Example Document-Feature Matrix Including Class

Art_ID	the	and	policemen	protesters	attack	attacking	pelted	...	Police Violence
1323742	65	40	0	1	0	0	0	...	TRUE
1302245	19	7	0	1	0	0	0	...	FALSE
670056	62	24	0	2	0	0	0	...	TRUE
657629	39	2	3	1	2	1	1	...	FALSE
1181330	32	12	0	0	1	0	0	...	FALSE

* The column '...' stands for omitted columns of which there are many.

last column, which is missing in the uncoded data so far or removed from the test data for validation. It intuitively makes sense that words like "attack" and "pelted" should be good indicators for framing focusing on trouble. Yet other words, like "the" and "and" are probably less relevant in deciding if a frame is present in an article or not. The difference between "attack" and "attacking" is also not relevant here.

Mostly the advice is, therefore, to eliminate some of the redundancy and unnecessary diversity of language through a number of preprocessing steps (Grimmer and Stewart, 2013; Jurafsky and Martin, 2020; Manning et al., 2008). In theory, reducing the set of features can have two advantages for SML: it reduces the size of data that needs to be evaluated for training and prediction, making the process more efficient; and it can actually improve accuracy of a classifier as a model can only predict classes based on the text features that were present in the training data.[35] However, there is no agreement which specific steps are suitable for text classification. In a review of the most often cited articles which employ unsupervised learning, Denny and Spirling (2018) find that there is no standard but, in fact, substantial differences in the choices for preprocessing steps. Usually though it is not discussed why a preprocessing step was performed or not. In the absence of a consensus on what preprocessing steps should be included for machine learning, it is difficult to decide what is most appropriate here.

Instead of emulating previous choices I, therefore, follow the suggestion by Denny and Spirling (2018), which is still relatively new in the field, to determine the optimal

[35] For example, if the test data contains the word "attacking", yet only the word "attack" was present in the training data, "attack" is ignored during prediction. Merging the two forms of the same word eliminates this problem.

preprocessing chain empirically. Specifically, I use all combinations of eight commonly employed preprocessing steps to prepare the articles for machine learning classification. Following the notation used by Denny and Spirling (2018) these are: removing *punctuation* from the text (**P**); removing *numbers* and other symbols (**N**); *lowercasing* (**L**) all words; *stemming* words (**S**), which is the process of finding common stems in order to merge several words into one (so that "attack" and "attacking" become the same feature); removal of so-called *stopwords*, which means words that do not convey much information, such as function words like "the" and "and" (**W**)[36]; removing *infrequently used terms*, which can not usually be used for machine learning anyway (if they appear in the test but not in the training set) (**I**)[37]; and including n-grams, which means to add word combinations to the data to capture cases in which the word order does play a role after all (**3**)[38]. Additionally, I included the step to weight features by their rarity in the document, specifically by using the *term frequency by inverse document frequency* (tf-idf) weighting (**T**) (Grimmer and Stewart, 2013; Jurafsky and Martin, 2020; Manning et al., 2008). This resulted in 256 (2^8) datasets as each combination of steps was tested. All preprocessing was done using *quanteda* (Benoit et al., 2018).[39]

To determine the optimal models and preprocessing chains for all seven frames, training portion of each of these 256 datasets was used to train machine learning models. Specifically, I used nine different algorithms implemented in the R software environment: *multinomial naïve bayes*, *support vector machine* and *penalized logistic regression* from *quanteda.textmodels* (Benoit et al., 2020c). *LogitBoost* from *caTools* (Tuszynski, 2020). *Bagging* from the *ipred* package (Peters and Hothorn, 2019). *Random forest* from the *ranger* package (Wright and Ziegler, 2017). *Decision tree* from the *tree* package (Ripley, 2019). *Multilayer perceptron network model* from *quanteda.classifiers*

[36] The specific list of stopwords that was used was the *English* set from the R package *stopwords* (Benoit et al., 2020b).

[37] Specifically I discard words which appear in less than 1% of documents.

[38] I include uni-, bi- and trigrams. Common example of bigrams are the difference between "national defense" and "national debt". Both "defense" and "debt" effectively change their meaning when used together with "national".

[39] Code for this batch evaluation of preprocessing steps is available as an R package: github.com/JB-Gruber/smlhelper.

Table 6.11: Classification Performance of Frames (Top 3 Models)

Frame	Accuracy	Precision	Recall	F1	Algorithm	Preprocessing
Police Violence	0.885	0.892	0.943	0.917	textmodel_nb	P-L-I-T-3
Police Violence	0.865	0.850	0.971	0.907	textmodel_svm	P-N-W-I
Police Violence	0.865	0.850	0.971	0.907	maxent	P-N-I-T-3
Law & Order	0.885	0.882	1.000	0.938	textmodel_svm	P-L-S-W-I-T-3
Law & Order	0.885	0.882	1.000	0.938	textmodel_svm	P-N-L-S-I-T-3
Law & Order	0.885	0.882	1.000	0.938	textmodel_svm	P-L-S-I-T-3
Cause & Grievances	0.769	0.789	0.882	0.833	textmodel_svm	S
Cause & Grievances	0.750	0.744	0.941	0.831	textmodel_svm	P-N-3
Cause & Grievances	0.750	0.744	0.941	0.831	textmodel_svm	N-3
Righteous Struggle	0.885	0.880	1.000	0.936	textmodel_svm	I-T-3
Righteous Struggle	0.885	0.880	1.000	0.936	textmodel_svm	L-S-W-3
Righteous Struggle	0.885	0.880	1.000	0.936	textmodel_svm	P-N-S
Troublemakers	0.827	0.848	0.951	0.897	maxent	N-W-I
Troublemakers	0.827	0.848	0.951	0.897	maxent	W-I
Troublemakers	0.827	0.848	0.951	0.897	maxent	N-W
Decay of Morals	0.904	0.915	0.977	0.945	maxent	N-W-I-T-3
Decay of Morals	0.904	0.915	0.977	0.945	maxent	P-N-L-3
Decay of Morals	0.904	0.915	0.977	0.945	maxent	P-L-3
Nuisance	0.904	0.898	1.000	0.946	maxent	P-S-I
Nuisance	0.904	0.898	1.000	0.946	maxent	P-N
Nuisance	0.904	0.898	1.000	0.946	maxent	P

(Benoit et al., 2020a). And *maximum entropy* from the *maxent* package (Jurka and Tsuruoka, 2013).

Performance of these models was then validated by comparing the machine prediction against the classes determined through factor scores. In the case of SML, the validation process mirrors the approach to test reliability in manual content analysis, as explained in Section 4.2.4, with the only difference that one of the coders is a machine. Specifically, 10 percent of the 500 coded articles were randomly chosen as *test set*, while the remaining 450 articles were used to train the models.[40]

As explained in Section 4.2.4, the measures usually reported for validation of supervised machine learning are *precision* and *recall*, as well as their harmonic mean, referred to as F1-score (also see Grimmer and Stewart, 2013; Manning et al., 2008, pp. 142-144). Validating nine different algorithms with 256 different dataset for the seven different frames produced 16,128 ($9 * 256 * 7$) different validation results. Table 6.11 shows

[40] To establish that the random composition of training and test set does not bias the validation results, validation was repeated with several different seeds. While results did vary slightly, no problematic variation of the results was found.

only the three best results, according to the F1-score, for each frame for the sake of readability. Out of these best models, the poorest result was achieved for the frame *cause & grievances* where accuracy drops below 80% agreement and F1 is just slightly above 0.8. Clear guidelines on how to interpret the specific measures are scarce, as a justifiable general threshold does not exist. However, Graesser et al. (2011) suggest that their rule of thumb is that F-measure scores of 0.7 or higher are impressive, 0.30 to 0.69 are modest and 0.29 are deemed disappointing (p. 43). According to this rule of thumb, the validation results shown in Table 6.11 indicate *impressive* performance of the top models for all frames, although classification works better for some frames than for others. Furthermore, all models are on levels comparable with or surpassing performance accepted by previous research (e.g., Burscher et al., 2014; Kananovich, 2018; Sevenans and Vliegenthart, 2016).

Although a number of algorithms were included in the test, only *multinomial naïve bayes*, *support vector machine* and *maximum entropy* appear in Table 6.11. Other models often performed reasonably well, yet these three algorithms proved superior with the specific classification task at hand. Interestingly, the variation is greater in terms of the different preprocessing steps. For the first two frames, classification accuracy was best when the entire or almost the entire set of preprocessing steps was performed. For the frame *Cause & Grievances*, on the other hand, models performed better when the original document-feature matrix stayed basically untouched. This finding is somewhat surprising, as some of the steps, especially lowercasing and removing punctuation, are used in basically all SML studies. Yet the few systematic studies that have been done support the idea that not all preprocessing improves SML performance and for some problems, some preprocessing steps actually hurt model validity (Scharkow, 2013). Additionally it was tested if combining several classifiers into an *ensemble of classifiers*[41] would yield better results — which was not the case.

[41] The idea behind ensemble classifiers is to pool the results from multiple algorithms, usually by calculating the mean of probabilities the algorithms predict for a case being a certain class.

I used the configurations which were determined optimal to code each frame to reproduce manual coding of the 500 randomly selected articles on the remaining 26,996 articles. Specifically, for each of the seven frames I trained a model using the optimal preprocessing chain and the optimal algorithm determined above. For this, I used the entire manually coded dataset. The models were then used to predict the presence of frames in the remaining dataset.

6.3.2 Results

The overall results of the frame coding on the dataset of 27,496 newspaper articles about protest can be seen in Figure 6.8. As a key interest of this research is to link the framing analysis back to claims of the existence and prevalence of the *protest paradigm*, I use the grouping already made in Section 6.2.4 to colour bars. Since the *protest paradigm* is assumed to delegitimise, marginalise and/or demonise protest (McLeod and Hertog, 1999), I use "delegitimising" as a shorthand for the frames with close links to the *protest paradigm* and "legitimising" for frames which likely have the opposite effect.

Figure 6.8 makes apparent that there is a stark difference in how often the frames are used. The single most salient frame is *decay of morals*, which highlights that protesters are a fringe group who allegedly disturb the general political consensus by questioning legitimate authority (also see e.g., Di Cicco, 2010). It was found in 35% of the articles in the dataset. The *troublemakers* frame, which highlights the violence caused by protesters and the associated risks, is the third most prevalent; *nuisance* is on rank four; while *law & order*, which portrays the police as defending public order against a "rowdy" protest, is still found in 7% of articles, which is, however, the smallest number of articles between the frames. Taken together, the delegitimising frames are clearly in the majority in the coverage of protest in the UK.

The second single most prevalent frame, however, is a legitimising one: *cause & grievances* supports protesters by making their main issues or cause explicit and re-

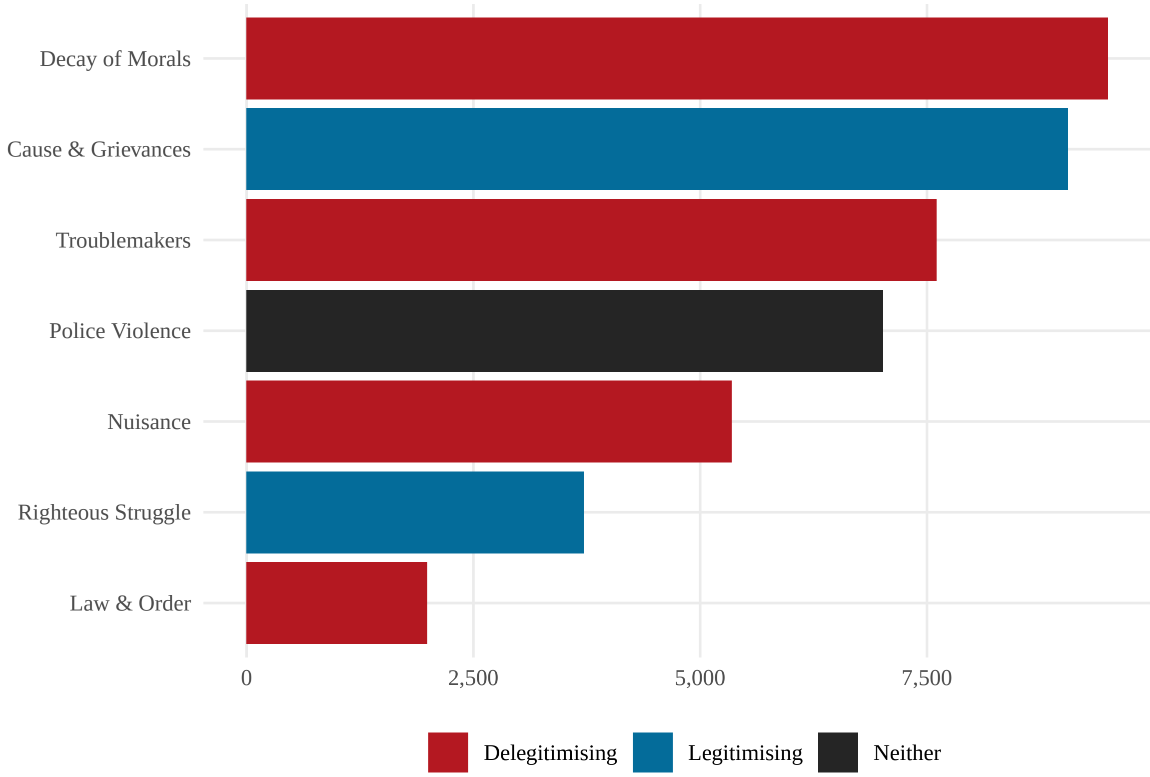


Figure 6.8: Number of Articles in Which Each Frame Is Present

latable. As *decay of morals*, it is present in roughly a third of all articles. *Righteous struggle*, which highlights the accomplishments of the protesters for a cause, is the second least prevalent, yet is still present in 14% of all articles. The one frame which can neither be deemed legitimising nor delegitimising sits at the middle of the prevalence ranking.

Figure 6.8 also shows that the total number between all frames surpasses the total size of the sample ($N = 27,496$). Again, this is possible as each article can contain several frames. At the same time, it also makes it tricky to evaluate if delegitimising or legitimising framing is more prevalent in the data: theoretically, it would be possible that, for example, several articles contain all four of the delegitimising frames, while each time a legitimising frame is counted, it is the sole frame in an article. To prevent a possible distortion, Figure 6.9 shows six of the seven frames grouped into delegitimising and legitimising coverage. Specifically, if *any* one of the frames *law &*

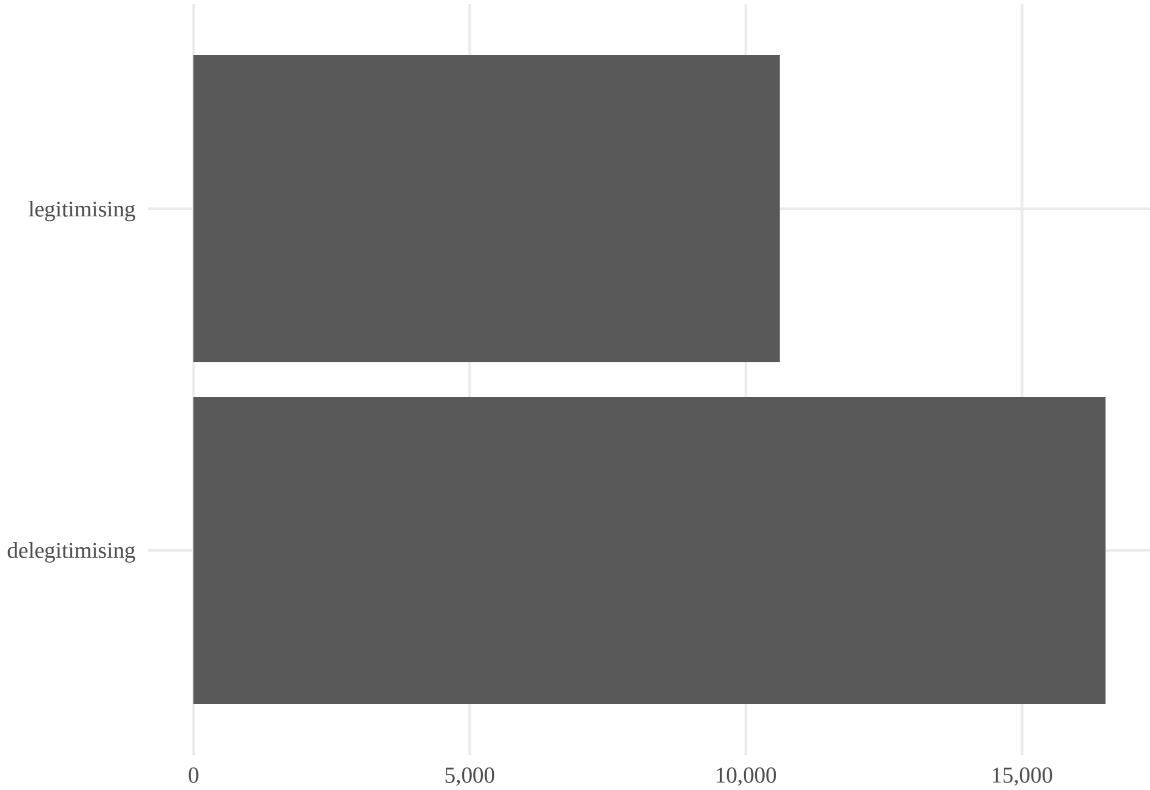


Figure 6.9: Number of Articles in Which Any (De-)Legitimising Frame Is Present

order, *troublemakers*, *decay of morals* and *nuisance* is present, the framing in an article is deemed delegitimising. If *cause & grievances* and *righteous struggle* are present, the framing in an article is deemed legitimising. Figure 6.9 confirms that delegitimising coverage is clearly present in a majority of the articles (60%) in the sample, whereas legitimising frames are present in 39% of the articles.

Additionally to overall prevalence, I am interested in the change over time. Figure 6.10 shows the number of articles using each frame in the full sample of 27,496 articles over the analysed time frame. Since the number of articles on protests varies between years (see Figure 5.3), the figure shows percentage of all articles that refer to a protest event in a year in which each frame was present. The numbers add up to more than 100% because many articles contain multiple frames (1.6 on average).

Again, as described in Section 6.2.4, four of the seven frames fall into the delegitimising spectrum: *law & order*, *troublemakers*, *decay of morals* and *nuisance*, while *cause &*

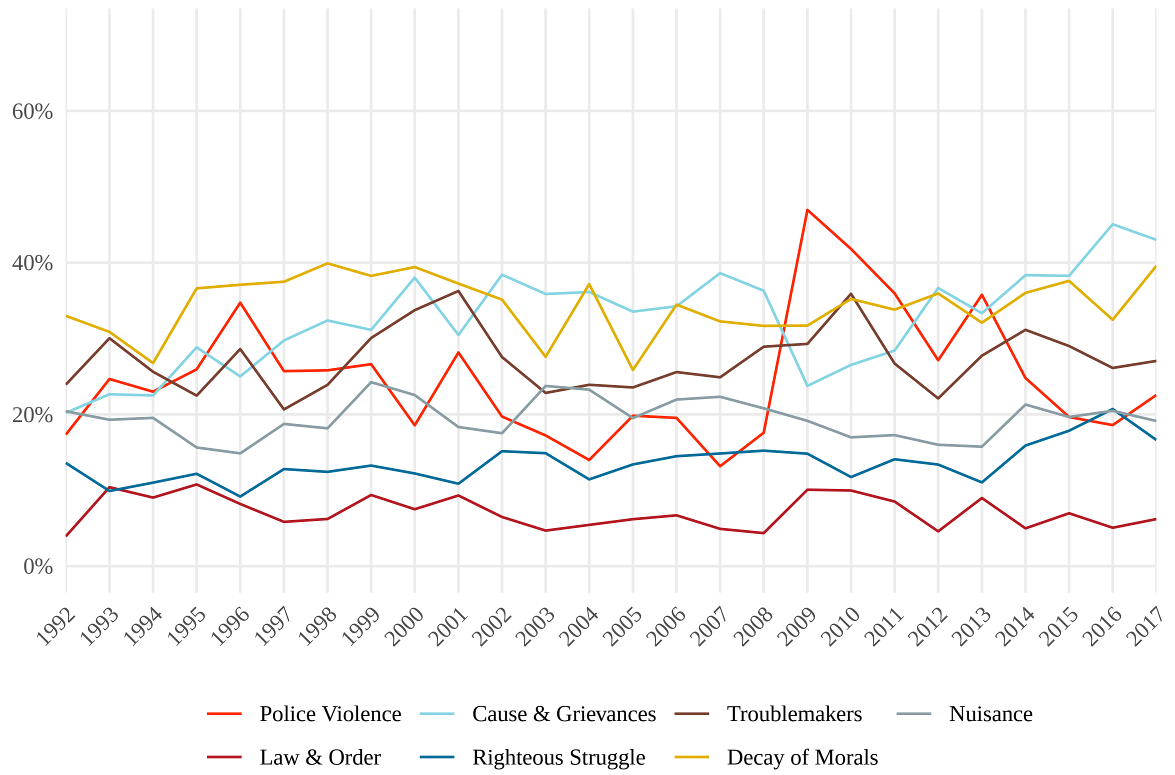


Figure 6.10: Percent of Total Articles per Year Containing a Frame

grievances and *righteous struggle* are deemed legitimising and *police violence* is considered neither legitimising nor delegitimising. As said, *law & order* is the least salient frame in the dataset. This is observed throughout the entire time, as it is present in between 10.8% and 3.9% of articles over the years. There does not seem any particular trend over time, with the last year being basically on the same level as the first year in the sample. For the frame *troublemakers*, we see some substantial peaks in 2000-2001 and in 2010. At both times police faced large and sometimes violent protests: in 2000 and 2001 there were the anti-capitalist May Day riots which made up the majority of reports using this frame, while in 2010 large student protests resulted in several violent incidents. Again, there does not seem to be a noteworthy general trend over the years. *Decay of morals*, as a relatively soft version of the *protest paradigm*, is the most often used frame in 12 out of the 26 years in the sample. However, there is no noteworthy trend over the years and the salience of the frame in 1992 and 2017 is quite similar.

The picture is similar for the *nuisance* frame, which stays present in around 20% of articles throughout the time period.

The two frames which were deemed legitimising are *cause & grievances* and *righteous struggle*. With *righteous struggle*, the pattern over time is similar to the one for the delegitimising frames: while there is an upward trend, it appears small. In most years, the frame plays a minor role, being just slightly more important than *law & order*. For the frame *cause & grievances*, however, a different picture emerges: from a starting point of 20.2% in 1992, the importance of the frame steadily increases over time until it reaches its peak in 2016, when the frame was present in 45.1% of all articles about protest. There is only one phase from 2009 to 2011 where the trend appears to be reversed. Yet the explanation is simple: Figure 6.10 shows relative values. As there is an explosive surge in the number of articles using either *police violence* or *troublemakers* in these years, *cause & grievances* looses in relative importance while the number of articles using the frame stays virtually the same between 2008 and 2012. In other words, *cause & grievances* is not used less often, there are just more articles employing the *police violence* and *troublemakers* frames.

For the remaining frame, *police violence*, which was deemed to not be clearly legitimising or delegitimising, we see the biggest variation in time, although again without a clear trend. As mentioned in Section 6.2.4, the prime example for this frame are the 2009 G20 summit protests in London during which, among other incidents of police brutality, the uninvolved Ian Tomlinson died after being beaten by police. The surge in articles employing the frame can be explained by this incident. A heap of articles during this time discussed police conduct in reports about the investigation and for many protest in the succeeding years. Again though, after interest had cooled down, the frame is back to its original salience in the last few years of the sample.

The impression that the share of most frames fluctuates but does not change substantially over time is formally confirmed through ordinary least squares (OLS) regression. I calculate one OLS model for each frame, using the year as independent variable

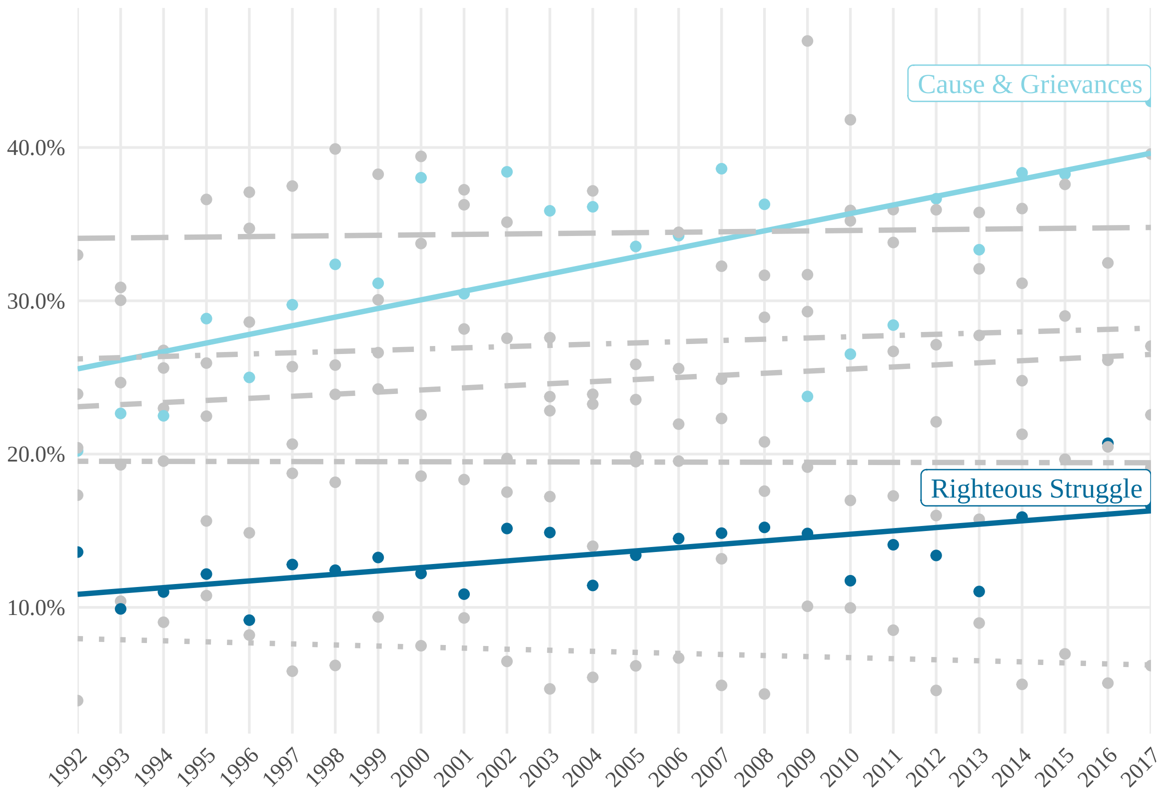


Figure 6.11: Percent of Total Articles per Year Containing a Frame, Significant Trends Highlighted

(starting with 1992 as year 0) and the percentage of total articles per year containing a frame as the dependant variable. Figure 6.11 shows the regression lines for each model in a single plot.[42] Non-significant relationships are greyed out while the significant ones keep the same colours for frames used in Figure 6.10. As can be seen, the salience of most frames stays constant and does not change significantly over time. Only the models for the two legitimising frames, *cause & grievances* and *righteous struggle*, show a statistically significant growth.

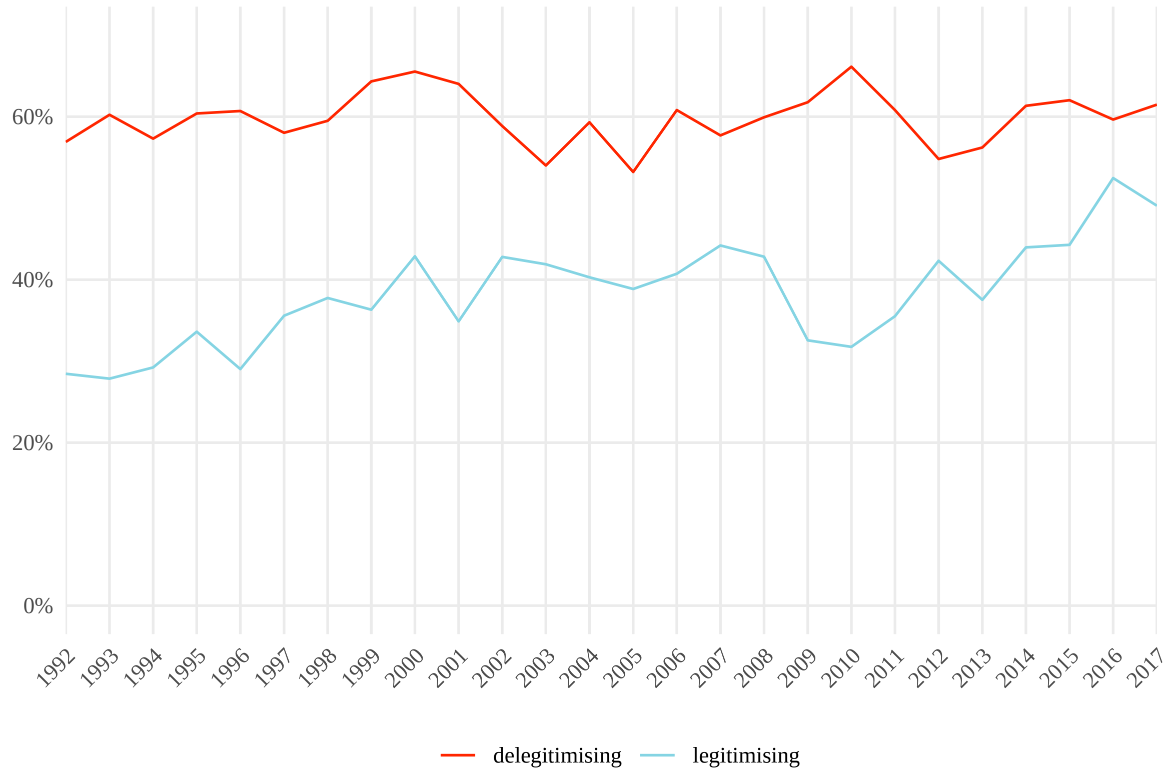


Figure 6.12: Percent of Total Articles per Year Containing (De-)Legitimising Coverage

Again, the reason why the salience of some frames grows while others stay constant is due to the fact that each article can contain several frames. I use the same grouping as in Figure 6.9 to rule out a possible distortion of prevalence of overall delegitimising or legitimising framing. Figure 6.12 shows that throughout the entire time, stories on protest events that contain *any* delegitimising coverage in the sense of the *protest paradigm* literature make up the majority. However, we also see that this dominance is

[42] For the full regression results see Appendix C.

slowly fading and almost half the articles contain legitimising aspects towards the final years as well. Again, I test if this trend is significant using OLS regression. The results show that the year is a significant predictor of the share of frames only in the model for legitimising frames (0.01*** (0.00)) but not for delegitimising frames. In other words, while legitimising framing gets more common over time, the share of delegitimising framing stays constant.

6.4 Discussion and Summary

This chapter has systematically analysed the framing of the newspaper coverage of domestic protest events in the UK. The aim was to assess how the news media frame the coverage of protest in eight selected newspapers from 1992-2017 (**RQ1**) and whether there has been a systematic change over time (**RQ2**). Since frames are an abstract variable, notoriously hard to identify and code in manual content analysis, coding them in such a broad and long-term study was no easy task. The approach taken was to split work into three steps: *manual coding*, *frame identification* and *frame coding*.

The manual coding demonstrated that splitting frames into smaller units makes content analysis more transparent, easier and yields more reliable results. The purpose of coding frame elements was to identify frames. The idea is that when certain frame elements are repeatedly used together, they constitute a frame as described by Entman (1993). The literature suggests to do this using cluster analysis (David et al., 2011; Matthes and Kohring, 2008). Cluster analysis groups together articles with similar selections of frame elements. This produces clusters in which each article is thought to use the same frame. Using cluster means, the frames can be identified as some frame elements are over- or under-represented in each cluster, making this the defining features of a frame. This led to an initial set of three broad frames: *Troublemakers*, *Struggle for (Just) Cause* and a category named *Mixed*. However, upon closer inspection, the class assigned to individual articles through cluster analysis often did not fit with manual assessment and omitted substantial nuances in the data. Theoretically, cluster analysis might be a poor choice for framing analysis due to the nature of the technique: each article can only be assigned one class, leading to the theoretically questionable situation of frames being mutually exclusive.

Dimension reduction was, therefore, performed again using factor analysis. The resulting factors showed overlap with the broad categories determined through the clustering, hence confirming the results trough triangulation. However, factor analysis retained a more nuanced picture of the coverage by identifying seven distinct frames: *police violence*, *law & order*, *cause & grievances*, *righteous struggle*, *troublemakers*, *decay of morals*, and *nuisance*. Each article might contain one, several or none of these frames, due to the nature of factor analysis as a mixed membership method. The comparison of the two methods showed that factor analysis should be regarded as the theoretical and empirically appropriate technique for identifying frames through dimension reduction. The seven frames identified through factor analysis are hence taken as the main frames in this study, which partially answers **RQ1**.

On a broad level, four of these seven frames can be described as delegitimising protesters efforts in the sense described in the *protest paradigm* literature: *law & order*, *troublemakers*, *decay of morals* and *nuisance*. *Cause & grievances* and *righteous struggle*, however, highlight the substantive content and background of protests or the positive aspect of it and are deemed legitimising. The remaining frame, *police violence*, can be seen as either delegitimising, as it might divert attention from the content of a protest to the details of the event. Or it could be seen as legitimising, since protesters are usually portrayed as the "good guys" who are attacked by police, which might lead to a positive view of them in the public. For the remaining analysis steps, it was decided to treat *police violence* as neither legitimising nor delegitimising.

Using factor scores, I determined the presence of the frames in the manually coded sample and reproduced this *frame coding* on the remaining articles. I find support for H1 (*Stories on protest events mainly use delegitimising framing as described in the protest paradigm literature*). An overall majority of reporting uses one or several frames with links to the marginalisation devices described in the *protest paradigm* literature. Broadly speaking, this thesis thus falls in line with previous studies which highlight the prevalence of reporting that follows a pre-defined recurring patterns of delegitimising story elements. However, the detailed framing analyses presented here adds nuance in several respects.

First, instead of only identifying the patterns of reporting in coverage, I quantify the salience of individual and grouped framings. Through this it becomes apparent that while delegitimising frames are present in the majority of the newspaper articles analysed, they are far from being completely dominant. The theory on the effects of framing presented in Section 3.1 can, therefore, help to better interpret the results: since a large part of reporting uses frames which highlight the messages of protest, it is reasonable to assume that this reporting drew attention to issues highlighted by protest, and possibly influence readers positively on how they regard the protest and the protesters. Contrary to previous assumptions, it is not at all clear then that coverage following the *protest paradigm* will lead to delegitimisation, marginalisation and demonization of protests (McLeod and Hertog, 1999), as alternative frames are reasonably available and accessible to the audience.

Second, I provide a more detailed picture of framing by evaluating the salience of individual frames rather than only the prevalence of broad ideal types. Specifically, I find that in the delegitimising frame category, *decay of morals* is the most prevalent, followed by *troublemakers*, *nuisance* and finally *law & order*. While these frames are all connected to the reporting devices described in the *protest paradigm* literature, they still carry different meanings and, presumably, strengths. *Decay of morals*, which suggests that protesters are a fringe group who disturb the political consensus, and *nuisance*, which suggests that protest is a bothersome interruption of everyday activities, might not be as delegitimising as *troublemakers* and *law & order*, which highlight violence. This means that while an overall majority of articles uses delegitimising framings, not all of these outright vilify protesters. Between the two legitimising frames, *cause & grievances* and *righteous struggle*, the first is widely more salient. Finally, I identified a frame which was neither delegitimising nor legitimising but was still present in a substantial portion of reporting. While it does not directly help to evaluate the hypotheses derived from the literature, it allows additional insight into the reporting.

A key aspect of this thesis is to scrutinise if the *protest paradigm* continues to be important for the reporting of protest. I find no support for H2a (*Delegitimising framing decreases in salience over time*) but accept H2b (*Legitimising framing increases in salience over time*). The results of the framing analysis largely support the existence and continued importance of a *protest paradigm*, which is thought to drive journalists to use delegitimising framing as the default in coverage of protest. However, the important contribution here is that it puts the salience of frames into a comparative perspective: the *protest paradigm* is apparently not the only force driving protest coverage — otherwise it would be more prevalent in terms of the observed framing of protests over the years. At no point in the time frame (1992-2017) did delegitimising coverage dominate the coverage of all protest. In fact, both of the legitimising frames are found to be present in a steadily growing share of protest reports — while none of the delegitimising frames gains over time.

Again the level of the individual frames provides some further insights. In 2009, after the death of Ian Tomlinson, the number of articles which scrutinised police conduct during protests exploded and only returned to the level of 2008 six years later. While it is not clear if this was a positive development for protesters, it means that the question of police conduct was more salient in reports about clashes during that time. Interestingly, the trend that the press increasingly portrays protest as a nuisance, which Di Cicco (2010) found for the United States, is entirely absent in the UK data.

Overall, the most important new insight is that the use of frames which highlight the goals and achievements of protests has grown over time. Although this was expected, the reasons are not so clear. Changes in the media landscape during the last decades have led to a situation in which newspapers have to fight for the attention of the audience who have an ever increasing choice of news and entertainment outlets. Perhaps journalists use legitimising frames more often today because they expect their readers to turn elsewhere for this information otherwise. At the same time, information on the goals and messages of protesters are more easily available than ever through the internet, especially social media sites, meaning that fewer resources are needed to include them in a report. Protest has also been normalised as a method of political participation over time, meaning the public must be expected to have more first-hand experience now than several decades ago. This means that they might be more critical of strictly delegitimising portrayals of events and less opposed to positive coverage of protest.

Nevertheless, coverage following the *protest paradigm* stays on a constant level in the time frame. This suggests that *news values*, especially the tendency to highlight the unusual, conflict and controversy, still drive the media to a large extent. However, in comparison to the grim conclusion about coverage of protest in the UK in works such as Halloran et al. (1970), Murdock (1973) and Glasgow University Media Group (1985), the press appears less hostile towards protesters in the 1990s and even less so in the succeeding decades. In other words, and answering **RQ2**, delegitimising framing of protest has not stopped and neither declined since 1992. But this is contrasted by an increasing level of legitimising framing of protest events, which has become almost as common in 2017. Now that these general trends of media framing of protest have been described, the next chapter asks *why* journalists choose the frames they choose for a protest.

Chapter 7

What Explains the Frames?

The last chapter has systematically explored how protest is framed. This chapter scrutinises *why* certain frames are applied. A core assumption of the *protest paradigm* literature is that certain news values, such as *conflict and controversy* or highlighting *the unusual*, work against protesters as they drive reporters to focus on the method of protest or the appearance of protesters over their message. As we have seen, this is true for a majority of reports about protest; however, a substantial number of articles still uses different frames. While the *protest paradigm* might be a default to a certain degree, reporters also regularly — and to an increasing degree — deviate from using delegitimising frames. The question which remains is: why? As argued in Section 3.4, a number of factors might drive reporting.

The purpose of this chapter is to answer the third and final research question and find out *which factors explain the choice of frames by the news media when covering domestic protest events?* To do that, the chapter is divided into three parts. Firstly, Section 7.1 discusses the specific variables and models I use. Secondly, Section 7.2 presents the analysis of the data and what it means for the hypotheses posed in Section 3.4. Like in the previous chapter, I group the individual identified frames into the broader categories of *legitimising* and *delegitimising* media framing to answer the theory-driven hypotheses. However, scrutinising only the use of aggregated frames would cast aside most of the additional detail added by using the careful inductive extraction of the most important frames, which was a key goal of this thesis. Section 7.3, therefore, focuses on zooming in on the level of the individual frames and, using the same modelling strategy as before, assessing under which circumstances these more specific frames are applied.

To briefly summarise the chapter findings, the analysis revealed a mixed fit with the theoretical expectations. Specifically, the idea that the *protest paradigm* drives journalists to use delegitimising frames more or less arbitrarily is supported. The only robust predictor for delegitimising frames is violence occurring during a protest. However, as in the previous chapter, this does not hold true for legitimising frames. *Labour* and *social issue* protests are given precedence and so are peaceful protests, during which no arrests are made. Additionally, broadsheet newspapers use legitimising frames more often and legitimising framing became more likely in recent years. Zooming in on the level of the individual frames provides further detail and shows, among other things, that different frames are employed for different kinds of protests and at different times. Especially within the delegitimising category, I find that some effects cancel each other out as, for example, the *law & order* and *nuisance* frames are more likely to be used when arrests are made, yet the *decay of morals* frame is less likely to be used in these situations, meaning delegitimising framing remains on the same level overall. Interestingly, the effects on the use of legitimising and delegitimising frames are not diametrically opposed. That means just because some protests are less likely to be framed delegitimising does not make them more likely to be reported on with a legitimising frame.

7.1 Data Processing and Models

To answer the remaining research question and assess the hypotheses, the framing analysis from Chapter 6 is used as the dependant variables in several multilevel logistic regressions. Table 7.1 shows the variables used in these models, where the data came

Table 7.1: Model Variables

Variable	Source	Scale	Processed	Levels/Range
Dependent				
Frame Use	Analysis Result	Dichotomous	-	0, 1
Event-Level				
Protest Goal	MMP Data	Nominal	Recoded	political*, social-issue, labour, anti-war, police
Protester Violence	MMP Data	Dichotomous	-	0*, 1
State Repression	MMP Data	Dichotomous	Combined	0*, 1
Outlet-Level				
Newspaper Ideology	LexisNexis Meta	Dichotomous	Combined	Left*, Right
Newspaper Type	LexisNexis Meta	Dichotomous	Combined	Tabloid*, Broadsheet
Ideological Divide	LexisNexis Meta	Nominal	Own Coding	Congruence*, Conflict, Ambiguous
Time-Bound				
Days Since Start	MMP + LN Meta	Interval	Combined + Recoded	0–1
Year of Protest	MMP Data	Ratio	Recoded	0–1
Level 2				
Newspaper-Year	LexisNexis Meta	Nominal	-	Newspaper x 1992–2017

* Reference category

from, their measurement scales and different levels. As explained in Chapter 4.3.1, the data was obtained from two sources: for the variables related to the newspapers, the metadata from the articles downloaded from the *LexisNexis* newspaper archive was combined with insights from the literature on the UK newspaper market; for the independent variables related to the protests themselves, I employed a subset of the *Mass Mobilization Project* (MMP) data (Clark and Regan, 2019) containing protests in the United Kingdom from 1992 to the end of 2017.

Frame Use in Table 7.1 is a stand-in since I run one model for each of the two aggregated and seven individual frames. The resulting continuous factor scores indicate how strongly an article is affiliated with a frame. However, due to the large number of articles about protest, this step was only performed for a training set of 500 randomly chosen articles. The remaining articles were classified using machine learning algorithms. Since machine learning algorithms made for text work best when predicting categories, however, this meant that the factor scores had to be dichotomised — negative factor scores were interpreted to show an absence of the frame, positive factor scores to indicate the presence of a frame. The dependant variables, therefore,

Table 7.2: Recoded Protester Demands

Code	Recoded to
political behavior, process	anti-war (also based on protester identity)
labor wage dispute	labour
police brutality	police brutality
land farm issue	social-issue
price increases, tax policy	social-issue
social restriction	social-issue
political behavior, process	political
removal of politician	political

Table 7.3: Distribution of Protester Demands

Demand	n	of total
political	3656	64.8%
social-issue	1102	19.5%
labour	566	10.0%
anti-war	225	4.0%
police	90	1.6%

contain, for each article, the information if a frame is absent (0) or present (1) in a news item. As explained in Section 4.3.2, not all reports coded in the last section could be matched with the data that provided the rest of the variables. The independent variables, therefore, contain information about frame use in 5,639 articles.

Most independent variables in the raw data were either enriched or recoded before the main analysis. The first event-level variable, *Protest Goal*, originally contained eight different demands from protesters. If a protest had more than one demand, Clark and Regan (2019) coded multiple goals for the same protest accordingly. The specific goals coded in the MMP data were recoded to achieve a better fit with the literature reviewed in Section 3.4. Table 7.2 shows the codes used by Clark and Regan (2019) on the left and the respective codes used here on the right.[43] To obtain the *anti-war* coding, protester identity, which is another code used in the MMP data, was used in addition to the protest demand. Table 7.3 shows how often each demand was coded.

[43] Specifically, the categories used here are *anti-war*, *labour*, *police brutality*, *social-issue* and *political* protests.

The variable *violence* is a proxy for protesters' tactics. As described above, tactics are usually operationalised by measuring if protesters broke the law and if they engaged in violence. Clark and Regan (2019), however, only provide dichotomous data, indicating whether protesters engaged in violence against the state in any form (1) or not (0). For the subset of the data, 3,431 of the events were coded as peaceful while violence occurred during 2,208 protests ($\bar{x} = 0.39$).

As mentioned in Section 4.3.1, a protest's goal and violence are prone to be affected by endogeneity, as Clark and Regan (2019) coded their data based on news reports, which is also the basis for the dependent variable here. It needs to be noted that the reports used for coding were removed before calculating the models.[44] Nevertheless, the results regarding the variables *protest goal* and *violence* should be seen with some caution.

The variable *state repression of peaceful protest* was constructed to reflect the occurrence of a specific case outlined by Wasow (2020): when the state uses repression against peaceful protesters. In this case, the expectation is that there will be more legitimising framing in the coverage of a protest event. The variable *protester violence* was used again to assess the second condition for this case. For the first condition, the variable *stateresponse* in the MMP dataset was employed. Of the five coded responses, "arrests" or "beatings" were interpreted as being repressive. When one of these two was employed by the state during peaceful protests, *state repression of peaceful protest* was coded as 1, otherwise, it was 0 ($\bar{x} = 0.46$). Figure 7.1 shows the percentage of violent and peaceful protests facing different kinds of state responses in the subset of the MMP data. One problem becomes apparent in this: the data for the UK did not contain a single case in which beatings occurred during peaceful protests. Therefore, in this case, *state repression of peaceful protest* reflects arrests of peaceful protesters.

[44] Specifically, the MMP data contains a text snippet from the article(s) on which the coding was based. These articles were removed from the database. Additional robustness checks were also performed and are described later on.

This is specific to the UK, data whereas beatings of peaceful protesters were recorded in other countries, including other countries in Western Europe.

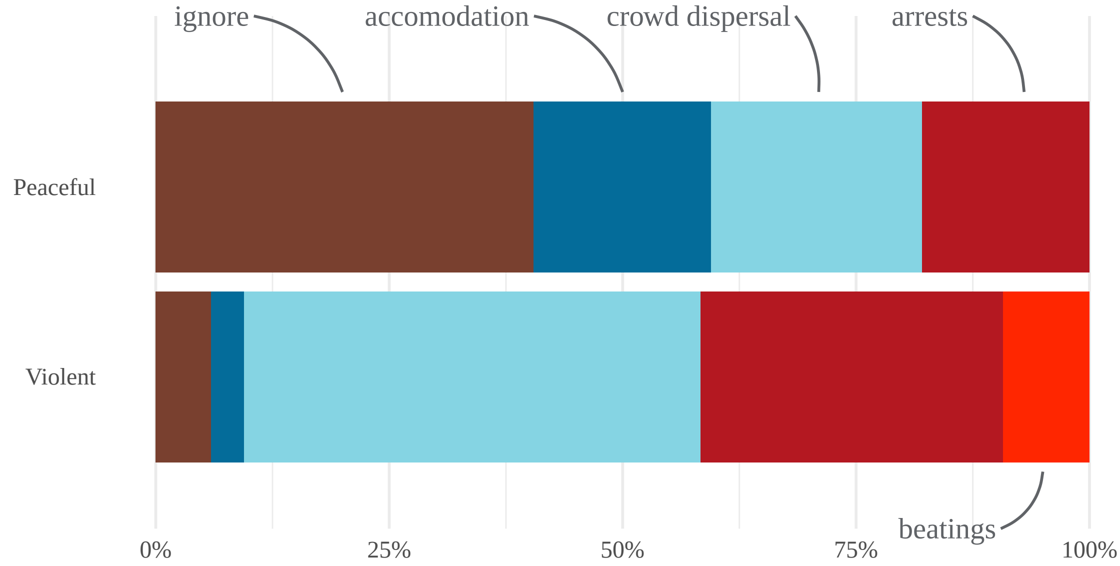


Figure 7.1: Percent of Violent and Peaceful Protests Facing Different Kinds of State Responses

The three outlet-level variables were coded based on the literature on the characteristics of the UK newspapers. Specifically, newspaper outlets in the UK are divided into right-leaning (i.e., closer to the Conservative party) and left-leaning (i.e., closer to the Labour party), as well as between tabloids and broadsheets. The newspaper outlets included in the sample were coded as noted in Table 5.1.

For the variable *ideological divide*, additional manual coding was necessary: An ideological divide is present when the newspaper ideology is right-leaning and the protest ideology is coded as left — or the other way around. The newspaper ideology is known; however, I had to code the ideology of the protest myself based on the information in the MMP dataset and in other articles about an event.[45] Note, that since “right”- and “left”-leaning are concepts which are often disputed, I tried to assess whether a protest aligns closer with the main UK right-wing or left-wing party (i.e., closer to Labour or Conservative politics), as this is how British newspapers are usually divided (Kuhn, 2007). Coding showed that between 1992 and 2017, 71% ($n = 4,021$) of protest

[45] See Appendix D for the specific coding.

events in the UK have been left-leaning, while 22% ($n = 1,233$) have been right-leaning. When the ideology of the protest and newspaper were the same, the *ideological divide* was coded as "congruence", when they differed, it was coded as "conflict". For 7% of protests, the ideological leaning could not be determined. This category comprises a heterogeneous group of events: ex-pat protests about issues abroad, protests against sectarian violence, protest against business practices of a company, protests demanding pay rises for police officers, prison guards and other groups who are formally not permitted to go on strikes, to name a few. For cases in which the protest ideology could not be coded, I added the category "ideological ambiguity", regardless of the newspaper ideology.

Finally, I used two *time-bound* variables: *days from start* and *year of protest*. The MMP dataset provides the first day of a protest. *Days from start* is then simply the number of days between the day a protest started and the day when an article was published. Since newspapers sometimes report about the preparations for a protest event before the actual date, the variable can have negative values. *Year of protest* indicates when a protest started, rather than the year in which reports about it were published. Both variables related to time were normalised by scaling them between 0 and 1 to make effect sizes comparable to the other variables — which all already range between 0 and 1.

As mentioned above, I fitted multilevel logistic regression models to establish the influence of the different factors on reporting. Specifically, I used the *lme4* package in R (Bates et al., 2015). The choice for this kind of model was made for two reasons: first, the dependent variables for all models are binary. A frame is either present (1) or absent (0), making logistic regression the appropriate tool. Second, in scenarios with panel data (i.e., with nested observations) the assumption of independence between cases, which is key for methods like OLS regression, is usually not met (Field et al., 2012). Specifically, it is likely that articles published in the same newspaper and in the same year are similar in content as it is shaped by specific cultural cues, routines, policies, structures and incentives within one news organisations in a given year. In such cases, a multilevel model (MLM) is necessary to control for correlated error (Gelman and Hill, 2006). Multilevel modelling aims at disentangling effects on different levels. If this is not done, standard errors are underestimated, which means the model might incorrectly show significant correlations where non exist (Ziller, 2018). It was considered to use newspapers as the second level variable as unobserved and unobservable heterogeneity is most likely present between outlets. However, the number of newspapers in the dataset — eight — is considered problematically low for fitting a valid multilevel model (e.g., Bryan and Jenkins, 2016). Therefore, I used newspaper-year as the Level 2 variable ($n = 180$). That means that *The Guardian*, for example, is treated as a different second-level case in 1992, 1993, 1994 and so on. This makes sense as the last chapter has shown that there is a difference in reporting between all newspaper over time.

Formally, the models for all frames have the same equation:

$$\begin{aligned}\log\left[\frac{P(\text{frame x} = \text{TRUE})}{1 - P(\text{frame x} = \text{TRUE})}\right] = \alpha_j + \beta_1(\text{goal}_{\text{anti-war}})_j + \\ \beta_2(\text{goal}_{\text{labour\_protests}})_j + \beta_3(\text{goal}_{\text{police}})_j + \\ \beta_4(\text{goal}_{\text{social-issue\_protests}})_j + \beta_5(\text{protesterviolence})_j + \\ \beta_6(\text{staterepression\_peaceful})_j + \beta_7(\text{np\_ideology}_{\text{right}})_j + \\ \beta_8(\text{np\_type}_{\text{tabloid}})_j + \beta_9(\text{ideo\_divide}_{\text{conflict}})_j + \\ \beta_{10}(\text{ideo\_divide}_{\text{ambiguous}})_j + \beta_{11}(\text{days\_since\_start})_j + \\ \beta_{12}(\text{year})_j + u_0(\text{Newspaper\_Year}) + \epsilon\end{aligned} \tag{7.1}$$

Frame x is again the stand-in for each of the tested frames. Level 1 in these models consists of newspaper articles nested in Level 2, which consists of newspaper-years. The β-values represent the level 1 variables described above. u_0(Newspaper_Year) is the error term for Level 2, which means it represents the unobserved heterogeneity on Level 2, thereby correcting the standard errors (Bell and Jones, 2015). Note, that the

Table 7.4: Comparison

	AIC NULL	AIC	BIC NULL	BIC	P
1. Trouble: police bad	7093.55	6459.20	7100.19	6472.48	0***
2. Trouble: police good	4177.17	4047.78	4183.81	4061.06	0***
3. Cause: officials bad	6943.54	6706.45	6950.18	6719.73	0***
4. Cause: protesters good	4397.27	4321.49	4403.91	4334.77	0***
5. Trouble: protesters bad	7216.56	6934.35	7223.20	6947.63	0***
6. Decay of morals	6801.45	6685.35	6808.08	6698.62	0***
12. Nuisance	5493.13	5406.17	5499.76	5419.45	0***
delegitimising	7480.63	7367.14	7487.27	7380.42	0***
legitimising	7425.75	7260.32	7432.39	7273.60	0***

[*] $***$ p $<$ 0.001; $**$ p<p0.01; $*$ p $<$ 0.05; + $<$ 0.1

models employ random, rather than fixed effects. Fixed effect (FE) approaches are more commonly used in time-series analysis in political science, yet, as Bell and Jones (2015) argue, random effects approaches are nearly always preferable. This is because unlike RE models, FE solutions discard a large amount of important information by controlling out all second level variance. Proponents of FE models believe that this is a justifiable trade-off, as RE models would risk suffering from *heterogeneity bias* as they are based on the assumption that u_0 does not correlate with any of the covariates in the model. If this was true, it would bias coefficients. However, Bell and Jones (2015) show through simulations that in practice, RE models perform at least as well as FE models while retaining the information about the level-2 level variance.

As a first step in multilevel modelling process, I assessed whether adding the newspaper level increases or decreases model fit. To compare models, the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) are calculated and compared using Analysis Of Variance (ANOVA) tests. The lower the AIC and BIC, the better the model fits the data. Table 7.4 shows that both AIC and BIC are always considerably lower when including the second level into the modelling. The p-values suggest that including the random effect improves all models significantly. Based on these unambiguous results I decided to use the model formalised above for the analysis of all frames.

As mentioned above, endogeneity is a problem for the analysis results shown below. Do the factors assessed below influence reporting or does the MMP data itself pick up trends in reporting, such as a tendency to focus on violence? Coding of the MMP data was based on different newspapers (Clark and Regan, 2019). The only direct overlap between the pool of newspaper Clark and Regan (2019) used for coding and the newspaper corpus employed here was *The Times*. Articles used for coding were therefore removed from the data before the analysis. To test the robustness of the results below, all models were run again after removing all articles published in *The Times* ($n = 1,123$). This did not cause any substantial changes (see Appendix E).

7.2 Hypotheses Testing

The objective of this section is to test hypotheses *H3a-m* set out in Section 3.4. Table 7.5 shows results from the two multilevel logistic models which assess the relationships between usage of de-/legitimising frames and the independent variables. The aggregated frames are an approximation made to discuss the hypotheses directly. As described above, an article is 1 on the dependent variable *delegitimising frames* if one or several of the frames *law & order*, *troublemakers*, *decay of morals* and *nuisance* are present and 0 if none of these frames is present; in the second model in Table 7.5, the dependent variable *legitimising frames* takes the value 1 if *cause & grievances*, *righteous struggle* or both are present in an article and 0 if none of these frames is present. The models discussed in the following section test the relationships between the usage of each of these individual frames and the independent variables. But as mentioned before, these frames were identified inductively and mostly had no direct equivalent in the literature. Discussing them is, therefore, done mostly in an explorative way, linking to the previous knowledge where possible. This and the next section discuss results following the same structure as Section 3.4: the seven independent variables are discussed one after another, grouped into event-level, outlet-level and time-bound factors.

	Delegitimising Frames	Legitimising Frames
(Intercept)	0.415^{***}	-0.428^{***}
	(0.123)	(0.124)
Event-Level Factors		
goal: anti-war	-0.140	-0.026
	(0.160)	(0.170)
goal: labour protests	0.059	0.278^{**}
	(0.105)	(0.105)
goal: police	0.277	-0.390
	(0.248)	(0.260)
goal: social-issue	-0.126	0.693^{***}
	(0.086)	(0.086)
violent protest	0.426^{***}	-0.474^{***}
	(0.074)	(0.075)
repression of peaceful p.	0.073	-0.577^{***}
	(0.093)	(0.097)
Outlet-Level Factors		
right-wing	0.056	-0.083
	(0.101)	(0.101)
tabloid newspaper	-0.066	-0.332^{**}
	(0.104)	(0.106)
ideological divide: conflict	0.018	-0.134
	(0.082)	(0.082)
ideological divide: ambiguous	0.229^{+}	0.481^{***}
	(0.134)	(0.133)
Time Bound Factors		
days from start	-0.563^{*}	-0.041
	(0.240)	(0.251)
year of protest	-0.083	0.349^{*}
	(0.170)	(0.170)
AIC	7337.030	7076.520
BIC	7429.954	7169.445
Log Likelihood	-3654.515	-3524.260
Num. obs.	5639	5639
Num. groups: np_year	180	180
Var: np_year (Intercept)	0.200	0.193

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{+}p < 0.1$

Table 7.5: Regression Results for Aggregated Frames

I recognise that while Table 7.5 is useful to assess the significance and direction of the impact the independent variables have on the outcome, the values of coefficient estimates are less useful. In ordinary least squares (OLS) regressions, estimated coefficients can be interpreted as marginal effects, meaning that an increase of 1 in the independent variable will result in the dependent variable to change by the value of the estimated coefficient. In logistic regressions, coefficients are less meaningful and can usually not be interpreted at all without further transformation (Leeper, 2021a; Long, 1997). Marginal effects are also not directly useful as they depend on the specific position of a data point on the non-linear slope of the regression surface.[46] Therefore, I additionally provide Average Marginal Effects (AME) in Figure 7.2. Holding all other covariate values constant, AMEs express here by how much the probability of using a frame increases or decreases *on average*, if a dependent variable changes by 1 in their respective units. For example, the AME of *violent protest* in the model for delegitimising frames is 0.10. This means that if a protest is violent instead of non-violent, that is the variable changes from 0 to 1, the probability that coverage contains a delegitimising frame increases on average by 10%. AMEs allow to easily express and interpret the influence of each covariate on the outcome. For the sake of clarity, Figure 7.2 contains only significant effects from the two models in Table 7.5. I calculate AMEs using the *margins* package in R (Leeper, 2021b).

7.2.1 Event-Level Factors

Protest Goals. Many scholars have highlighted the importance of goals or causes voiced by protesters as a key factor shaping the coverage these events receive. Often, the underlying assumption is that journalists tend to prefer certain protest issues which conform to a perceived status quo for a number of reasons, including ideology, news values and organisational routines (e.g., Boyle et al., 2012; McLeod and Hertog, 1999). The most recent of these studies, especially Kilgo and Harlow (2019), reject the

[46] For example, the marginal effect for a change in the variable "violent protest" from 0 to 1 will be different in each year of the data, even while holding all other variables constant.

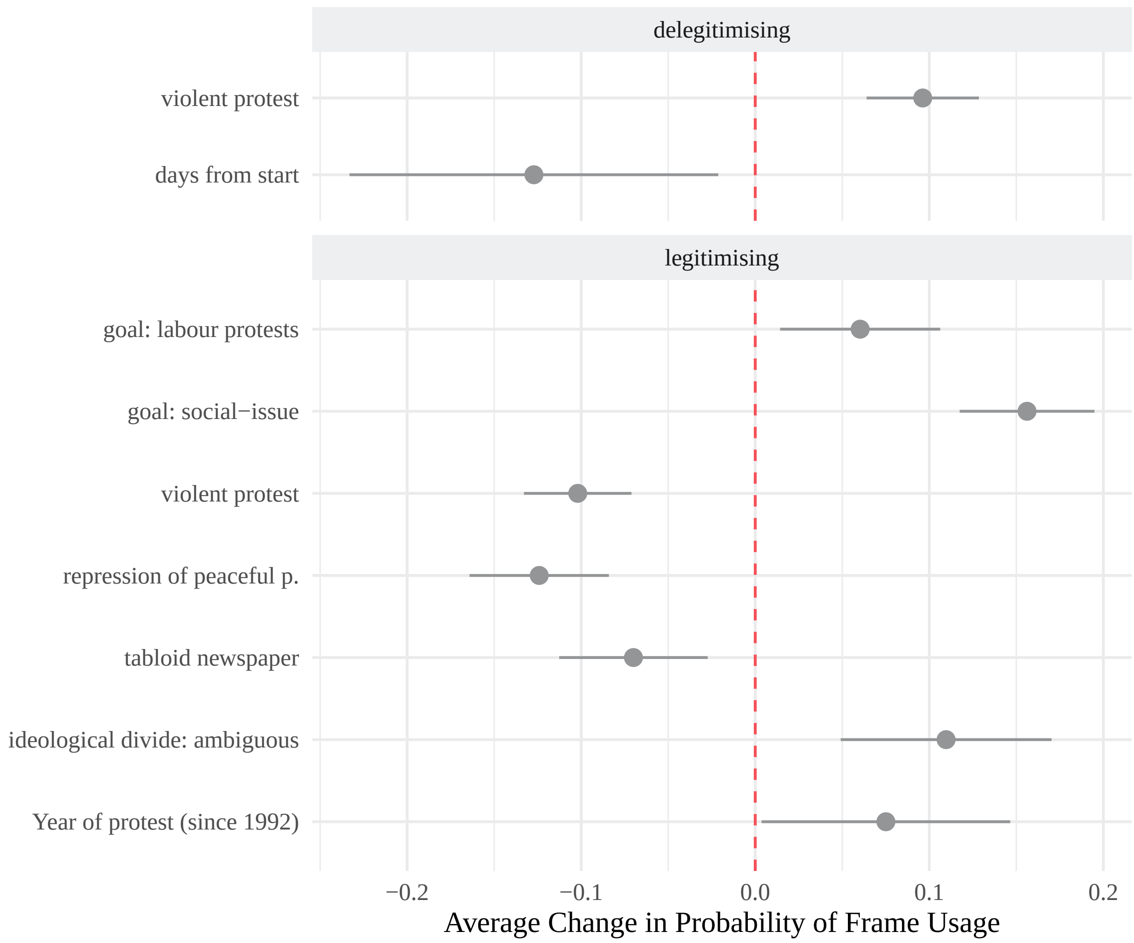


Figure 7.2: Average Marginal Effects for Significant Variables in Aggregated Frame Models

distinction of goals between support or threat to the status quo though as it is difficult to operationalise. Instead, they propose that there is a *hierarchy of social struggle* that determines which issues are legitimised and which are delegitimised in the coverage.

Looking at protest goals in Table 7.5, it is important to note that the results are in comparison to the reference category *political* protests, which was coded most often. For H3a (*The goal of a protest determines how likely delegitimising framing is to be used, following a hierarchy of social struggle that is ordered from highest to lowest probability: political, war, worker, police and social issues protests*), Table 7.5 shows that none of the goals has a significant effect on the probability that any delegitimising frame is used. This suggests that coverage of none of the goals is systematically different from the coverage of the reference category *political protest*. In other words, delegitimising frames are used for all goals similarly, which means that there is no hierarchy of social struggle in UK newspaper coverage — which would be the case if certain topics are given precedence and legitimacy while others are delegitimised (Kilgo and Harlow, 2019). **H3a** is therefore **rejected**.

The picture is different for legitimising coverage, i.e., H3b (*The goal of a protest determines how likely legitimising framing is to be used, following a hierarchy of social struggle that is ordered from highest to lowest probability: social-issue, police, labour, anti-war and political protests*). *Labour* protests have a significantly higher probability to receive legitimising coverage than *political* protests (AME: 6%), while *social-issue* protests have an even higher probability compared to *labour* protests (AME: 16%). In other words, a protest for higher salaries among policemen, a typical *labour* issue, will have, on average, a 6 percentage points higher probability of receiving legitimising framing, compared to a protest for measures against climate change, a typical issue of *political protests*. A *social issue*, such as a protest against tuition fees, is even more likely to be framed in a legitimising way. Specifically, its probability is, on average, 16 percentage points above the one for *political protests*. However, *anti-war* protests and *police* protests are not significantly different from the reference category.

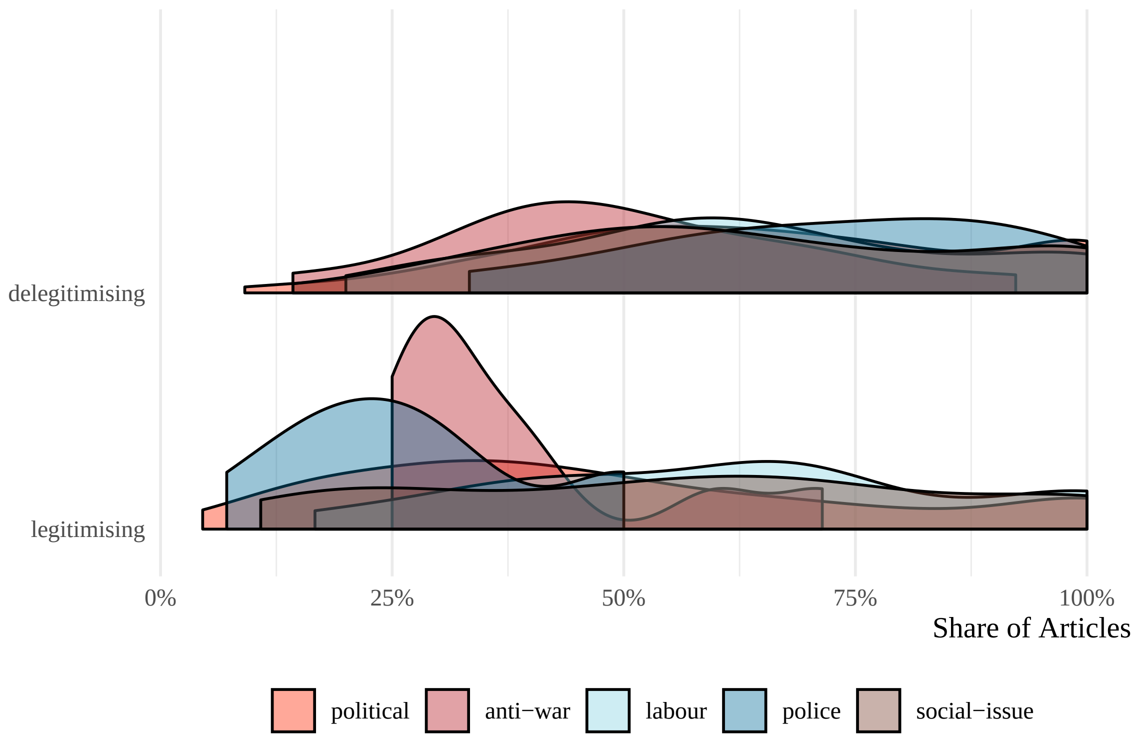


Figure 7.3: Presence of Frame in Percent of Articles About Protests with Different Goals

What does that mean? First, **H3b** can be partially accepted: protests with certain goals are more likely to be covered with a legitimising frame than others. However, what is the specific hierarchy of social struggle in the UK? To find this out I use Figure 7.3, which shows a ridgeline density visualization. This type of visualisation is based on density estimates, which show the distribution of data similar to a histogram. Several distributions are stacked on top of each other, each slightly transparent. This makes it easy to compare multiple distributions: If the ridges overlap, distributions are similar, if the peaks and valleys of the ridges differ, distributions are markedly different. In this case, it shows, for example, that the highest density (i.e., most values) for legitimising coverage and *anti-war* protest is at around 37%. That means that many events receive legitimising coverage in about 37% of the articles which cover it.

Figure 7.3 helps to show, in a visual way, why none of the different goals made a significant difference on whether a protest received coverage with one or several delegitimising frames: the distributions for delegitimising framing are almost identical between goals.

The two slight deviations from the overall trend are *anti-war* protests, for which the distribution is skewed more towards less delegitimising framing, and police protests, which are skewed towards more delegitimising framing.

For legitimising framing, the hierarchy of social struggles is as follows: protests against police brutality have the least chance to receive coverage using legitimising framing; followed by *political* protest, *anti-war* protests and labour protests; social-issue protests, finally, have the highest probability on average to be covered with a legitimising framing. There is, however, a level of uncertainty in the exact order of this hierarchy, as *political* and *anti-war* protests appear to have nearly identical probabilities to be framed in a legitimising frame. The distinction between protests against police brutality and the others is also only supported by the visual analysis in Figure 7.3 and is not significant in the regression presented in Table 7.5. Furthermore, the specific hierarchy which was expected did not emerge: as expected, *social-issue* protests are most likely to be covered with a legitimising frame. However the remaining expected order (*police*, *labour*, *anti-war* and *political* protests) was not found. Overall the data suggest that certain goals are given precedence compared to the others, yet findings about a hierarchy of social struggle depart from theoretical expectations. **H3b** is therefore rejected.

Tactics. As explained before, the relation between violent tactics and less favourable coverage of the protest is well established. Table 7.5 and Figure 7.2 confirm this connection. Violence has a significant effect in the expected directions in both the models for legitimising and delegitimising frames. When there is violence, the probability that a delegitimising frame is used increases by 10%. Conversely, when protesters remain peaceful, the likelihood for usage of legitimising frames increases by 10%. In other words, when protests become violent, it is a lot more likely that they will receive both more delegitimising and less legitimising coverage. Consequently, both H3c (*When protesters break laws or when they engage in violence, delegitimising framing is used more often*) and H3d (*When protesters are peaceful and obey the law, legitimising fram-*

ing is used more often) are accepted. Furthermore, some authors suggest that protests might use violent tactics to attract more media coverage (e.g., Boyle et al., 2012). The findings here suggest that the value of media exposure for a specific protest needs to be quite high for this trade-off to make sense.

While this outcome is far from a surprise, it has to be noted again that the variable *violence* is prone to suffer from endogeneity effects. As the coding is based on news reports, there is no objective way of knowing whether protesters engaged in violence — all we know is that it was *reported* that they did. To circumvent this problem, the original newspaper articles from which the variable was coded were removed from the data before the analysis. Additionally, the models were fitted again completely without articles from *The Times* (the only British newspaper which was used for some of the coding). This showed nearly identical estimates with 0.47*** (0.07) in the model for delegitimising frames and -0.50*** (0.08) in the model for legitimising frames, confirming the robustness of this finding.[47]

State Response. The notion that the state's response matters for protest coverage is based on findings by Wasow (2020), which suggest that in cases with violent reactions towards peaceful protest, coverage will be more legitimising. However, the findings here show the exact opposite effect: it appears that legitimising framing is *less* likely by 12% for events in which the state uses repressive tactics against peaceful protesters. This means that H3e (*When the state uses repressive tactics on peaceful protesters, legitimising framing is used more often*) has to be rejected. But why?

Looking again at Figure 7.1, it must be noted that the effect here is not driven by *violent* state responses to peaceful protest, as was the case in the study by Wasow (2020). There were no instances of violent state response towards peaceful protest in the UK data set.[48] Instead the variable *state repression of peaceful protest* reflects

[47] Full models in Appendix E.

[48] Unfortunately it was not possible to determine if this case never occurred or if it was just never coded in the MMP data for the United Kingdom. State violence was coded for peaceful protest elsewhere in the world though, including in liberal/Western democracies.

arrests of peaceful protesters here. The findings are nevertheless striking; it means that when protesters are arrested, this has a significantly negative impact on the probability of receiving coverage framed in a legitimising way, even if they remain peaceful.

The theoretical expectation for repression of peaceful protest does not include the case of delegitimising framing, which is why there was no hypothesis for it. What becomes apparent from Table 7.5 is that state repression, or more specifically in this case arrests, does, in fact, not significantly affect the probability that delegitimising frames are used.

7.2.2 Outlet-Level Factors

At the outlet level, the idea was, as explained in Section 3.4, that the specific routines, goals, policies and structures within news organisations shape content. The three factors which are taken into account are *newspaper ideology*, *ideological divide* and *newspaper type.*

Newspaper Ideology. To recap, there are different theories about why and how the ideology of a news organisation matters for protest coverage. Early studies like Chan and Lee (1984) found that left-leaning media were less hostile towards protest. This was picked up again by more recent studies, with the same result (e.g., Lee, 2014; Shahin et al., 2016). A conflicting theory suggests, however, that more than the specific ideology of a news organisation, what matters is if this ideology conflicts with the ideology of a protest or not (Weaver and Scacco, 2012). As right-leaning protest has become more widespread over time (Milne, 2005), the difference in overall reporting about protest between right- and left-leaning outlets could have been diminished.

Testing the two hypotheses which focus purely on *newspaper ideology*, we see that this variable shows no significant effect in either of the two models in Table 7.5. This means that both H3f (*Right media outlets use delegitimising framing more often than left media outlets*) and H3g (*Left media outlets use legitimising framing more often than right media outlets*) have to be rejected. The outcome is surprising, especially since the

UK case was specifically chosen because the ideological divide between outlets is more distinct than in the US or elsewhere (Hallin and Mancini, 2004; Kuhn, 2007) and has been suggested to shape UK coverage of protest as well (Gavin, 2007).

However, during the coding phase of this thesis, it became apparent that some protests were supported openly by right-leaning news outlets, such as the 2000 fuel protests or protests against fox hunting legislation introduced in 2005, while left newspapers seemed less supportive. So maybe ideological conflict, rather than newspaper ideology, matters for reporting? Again, the answer is no. Both models show no significant difference between ideological congruence and conflict between protesters and newspaper. Both H3h (*When there is a divide between the ideology of the protest's goal and the ideological leaning of the outlet, delegitimising framing is used more often*) and H3i (*When there is agreement between the ideology of the protest's goal and the ideological leaning of the outlet, legitimising framing is used more often*) are therefore rejected. This means that while there is anecdotal evidence that left-wing media are more critical of right-wing protest and vice versa, the phenomenon is not widespread enough to be picked up by systematic statistical analysis.

Surprisingly, the *ideological ambiguity* category, which means that a protest was neither left nor right, has a positive effect in both models — although only at the $p < 0.1$ level of significance in the model for delegitimising frames. In the model for legitimising frames, the ambiguity category shows a positive effect at an AME of 0.11. This is a substantial effect at the same magnitude as the factor *violence*. These protests are, as mentioned above, very heterogeneous, ranging from protest against business practices of a company to ex-pat protests about issues abroad. It is thus not straightforward to understand why they are more likely to be covered with legitimising frames *as well* as delegitimising frames. One possible explanation is that journalists feel less constrained to voice their opinion when an issue hasn't been discussed along party lines (yet). Another possible explanation would be that issues that are debated by parties need less explanation when they are picked up by protesters. If the grievance of a protest has not been discussed in public before and if a group that protests was previously unknown, articles might spend more time on interpreting an event.

Type of Newspaper. Moving on to the *type of newspaper*, we can see that this indicator only has an effect in the model for legitimising frames: it is significantly less likely for tabloid news media to use legitimising frames than it is for broadsheet outlets. Specifically, the probability that a legitimising frame is used is 7 percentage points lower on average when an article is published in a tabloid newspaper. For delegitimising frames, it seems that both types of media outlets are equally likely to use delegitimising frames. Hypothesis H3j (*Tabloid media outlets use delegitimising framing more often than Broadsheet outlets*) is therefore rejected, while H3k (*Broadsheet media outlets use legitimising framing more often than Tabloid media outlets*) is accepted.

7.2.3 Time-Bound Factors

The remaining two factors are connected to time. The model for delegitimising frames shows that as more days pass after the beginning of a protest, it becomes more likely that articles will contain a delegitimising frame. The AME shows that the effect is one of the largest at a -13%. However, remember that *days from start* was one of the two variables that were rescaled so that the minimum was 0 and the maximum was one. The AME here, therefore, displays the change in probability when going from the minimum of 7 days before, to the maximum of 81 days after the start of a protest.

Since the variable is continuous, we can also look at the model prediction directly in a plot. Figure 7.4 shows predicted probabilities for all articles in the dataset as points and a trend line as visual help to see the resulting patterns.[49] As becomes apparent in Figure 7.4, the impact of *days from start* on outcome does not follow a linear path.[50] Instead, we see the highest average probabilities that one of the delegitimising frames

[49] Technical note: commonly this kind of plot is shown with confidence intervals. Unfortunately though, the calculation of correct standard errors in predictions using multilevel models is not trivial and has not been implemented in the used lme4 package yet. See github.com/lme4/lme4/issues/147.

[50] The x-axis of Figure 7.4 was scaled back to the original values but is based on respective values between 0 and 1 as used in the model.

is used in the first week after a protest, with lower values before and after. Average probabilities decrease for a while but increase again after about 8 weeks — a time frame in which only a few high-profile and long-lasting protests still receive any coverage at all. It seems then that delegitimising frames are most compelling to journalists in the direct aftermath of an event. This confirms the expectation in H3l (*At the beginning of protest coverage of an event, news will be event-driven and hence are more likely to contain delegitimising framing*).

Note, however, that this does not at all mean that legitimising frames will be used more often as the protest develops. As each article can contain one or multiple delegitimising or legitimising frames or neither, the two models must be seen independently from each other. And indeed, the model for legitimising frames shows no significant effect for the factor *days from start*. Hypothesis H3m (*The more time passes between the start of a protest event and publication of an article about it, the more likely it is that the article contains legitimising coverage*) is therefore rejected.

As was already shown in Chapter 6, the year doesn't affect the probability of delegitimising frames but does increase the probability that legitimising frames being used.[51] The full model, therefore, confirms this previous finding as the year has a significant positive effect on the probability that legitimising framing is used, suggesting legitimising frames become more salient over time. Specifically, comparing the last year, 2017, in the dataset to the first year, 1992, legitimising framing has become 8% more likely on average.

7.3 Unpacking Frame Usage

In the previous section, the theoretical expectations towards what influences the usage of frames were tested. This was done using grouped dependent variables signifying if one of several delegitimising or legitimising frames were used. However, a key goal of

[51] Note though, that in Chapter 6 the year in which an article was published was used while here it is the year of the protest event itself — which is identical for most articles though. Note also that articles were excluded if they could not be matched with any event in the MMP data.

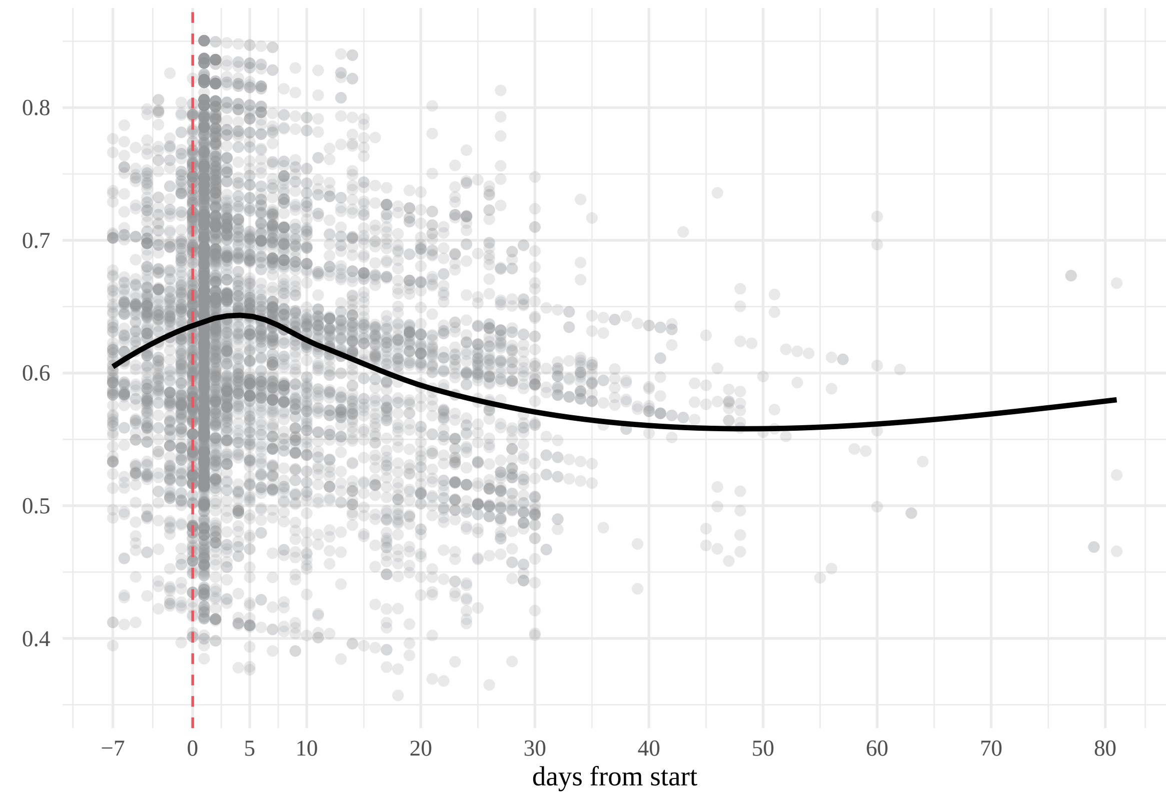


Figure 7.4: Predicted Probabilities of Using a Delegitimising Frame Before and After the Start of a Protest

this thesis was to assess what specific frames are used in reporting different protests. This section, therefore, discusses regression results for seven different models, each using a different frame as the dependent variable. The data and model specifications for all of these models were identical to what was discussed in Section 7.1. Discussion of the results follows the same structure as the previous section.

7.3.1 Event-Level Factors

Protest Goals. In the previous section, I concluded that none of the goals has a significant effect on the probability that any delegitimising frame is used. This is true when the usage of delegitimising frames is combined. However, on the level of the individual frames, we see several significant values in Table 7.6. *Nuisance* is used significantly more likely for *labour* protests than for the reference category (AME: 0.06). For protests against *police brutality*, the effects on the different frames actually cancel each other out: while it is significantly more likely that *law & order* (AME: 0.23) and *troublemakers* (AME: 0.12) are used in coverage of police protests, it is also significantly less likely that *decay of morals* (AME: -0.14) and/or *nuisance* (AME: -0.09) are used. In other words, some delegitimising frames are more likely to be used while others are less likely to be used, which, in sum, means that the probability that *any* delegitimising frame is used stays roughly the same.

It is interesting to see the probability that *law & order* is to be used, which highlights the positive impact police presence has, increases for protests against police brutality. However, the model for the frame which has the exact opposite meaning, *police violence*, shows an even higher AME of 37%. That means that protests against police brutality are divisive: they lead to more coverage highlighting claims of police brutality and providing background on them; but coverage will also defend police actions. The final goal, *social issues*, only significantly changes the outcome in the model for the *law & order* frame. Specifically, it is 4% less likely that the *law & order* frame is used for *social-issue* protests compared to *political* protests.

	Delegitimising				Legitimising		
	Law & Order	Troublemakers	Decay of Morals	Nuisance	Cause & Grievances	Righteous Struggle	Police Violence
(Intercept)	-2.498^{***}	-0.985^{***}	-0.862^{***}	-1.664^{***}	-0.960^{***}	-1.467^{***}	-2.016^{***}
	(0.194)	(0.142)	(0.150)	(0.139)	(0.145)	(0.148)	(0.183)
Event-Level Factors							
goal: anti-war	-0.010	-0.105	0.039	-0.066	0.067	-0.081	-0.114
	(0.294)	(0.188)	(0.186)	(0.189)	(0.179)	(0.247)	(0.206)
goal: labour protests	-0.021	-0.169	0.030	0.347^{**}	0.306^{**}	0.067	-0.319^{*}
	(0.171)	(0.117)	(0.110)	(0.115)	(0.112)	(0.142)	(0.132)
goal: police	1.528^{***}	0.566^{*}	-0.916^{**}	-0.797^{*}	-0.318	-0.423	1.949^{***}
	(0.258)	(0.235)	(0.326)	(0.382)	(0.283)	(0.366)	(0.292)
goal: social-issue	-0.507^{***}	-0.093	0.079	0.058	0.916^{***}	-0.240^{+}	-0.657^{***}
	(0.140)	(0.092)	(0.092)	(0.102)	(0.091)	(0.123)	(0.109)
violent protest	1.355^{***}	0.779^{***}	-0.177^{*}	-0.379^{***}	-0.609^{***}	-0.062	1.567^{***}
	(0.120)	(0.078)	(0.080)	(0.092)	(0.081)	(0.099)	(0.087)
repression of peaceful p.	0.462^{**}	-0.131	-0.455^{***}	0.317^{**}	-0.559^{***}	-0.553^{***}	0.325^{**}
	(0.166)	(0.105)	(0.105)	(0.107)	(0.103)	(0.146)	(0.118)
Outlet-Level Factors							
right-wing	-0.262^{+}	-0.020	0.145	0.155	-0.052	-0.320^{**}	-0.216
	(0.156)	(0.115)	(0.123)	(0.113)	(0.117)	(0.122)	(0.151)
tabloid newspaper	0.212	0.159	-0.097	-0.093	-0.228^{+}	-0.448^{***}	0.342^{*}
	(0.154)	(0.118)	(0.129)	(0.116)	(0.123)	(0.132)	(0.153)
ideological divide: conflict	0.259^{+}	0.171^{+}	-0.173^{+}	0.187^{+}	-0.172^{+}	-0.063	0.267^{*}
	(0.132)	(0.088)	(0.090)	(0.097)	(0.089)	(0.112)	(0.107)
ideological divide: ambiguous	0.717^{**}	0.696^{***}	-0.034	0.337^{*}	0.412^{**}	0.259	0.094
	(0.223)	(0.140)	(0.143)	(0.153)	(0.138)	(0.170)	(0.179)
Time Bound Factors							
days from start	-0.947^{*}	-1.426^{***}	0.680^{**}	-0.224	0.222	-0.659^{+}	-0.233
	(0.407)	(0.272)	(0.256)	(0.302)	(0.267)	(0.360)	(0.288)
year of protest	-0.581^{*}	0.050	-0.159	0.065	0.627^{**}	0.037	0.715^{**}
	(0.257)	(0.194)	(0.209)	(0.189)	(0.200)	(0.203)	(0.251)
AIC	3811.033	6745.761	6663.276	5349.929	6472.835	4290.523	5854.584
BIC	3903.957	6838.686	6756.200	5442.853	6565.760	4383.448	5947.508
Log Likelihood	−1891.516	−3358.881	−3317.638	−2660.964	−3222.418	−2131.262	−2913.292
Num. obs.	5639	5639	5639	5639	5639	5639	5639
Num. groups: np_year	180	180	180	180	180	180	180
Var: np_year (Intercept)	0.378	0.289	0.357	0.193	0.311	0.183	0.559

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{+}p < 0.1$

Table 7.6: Regression Results for Individual Frames

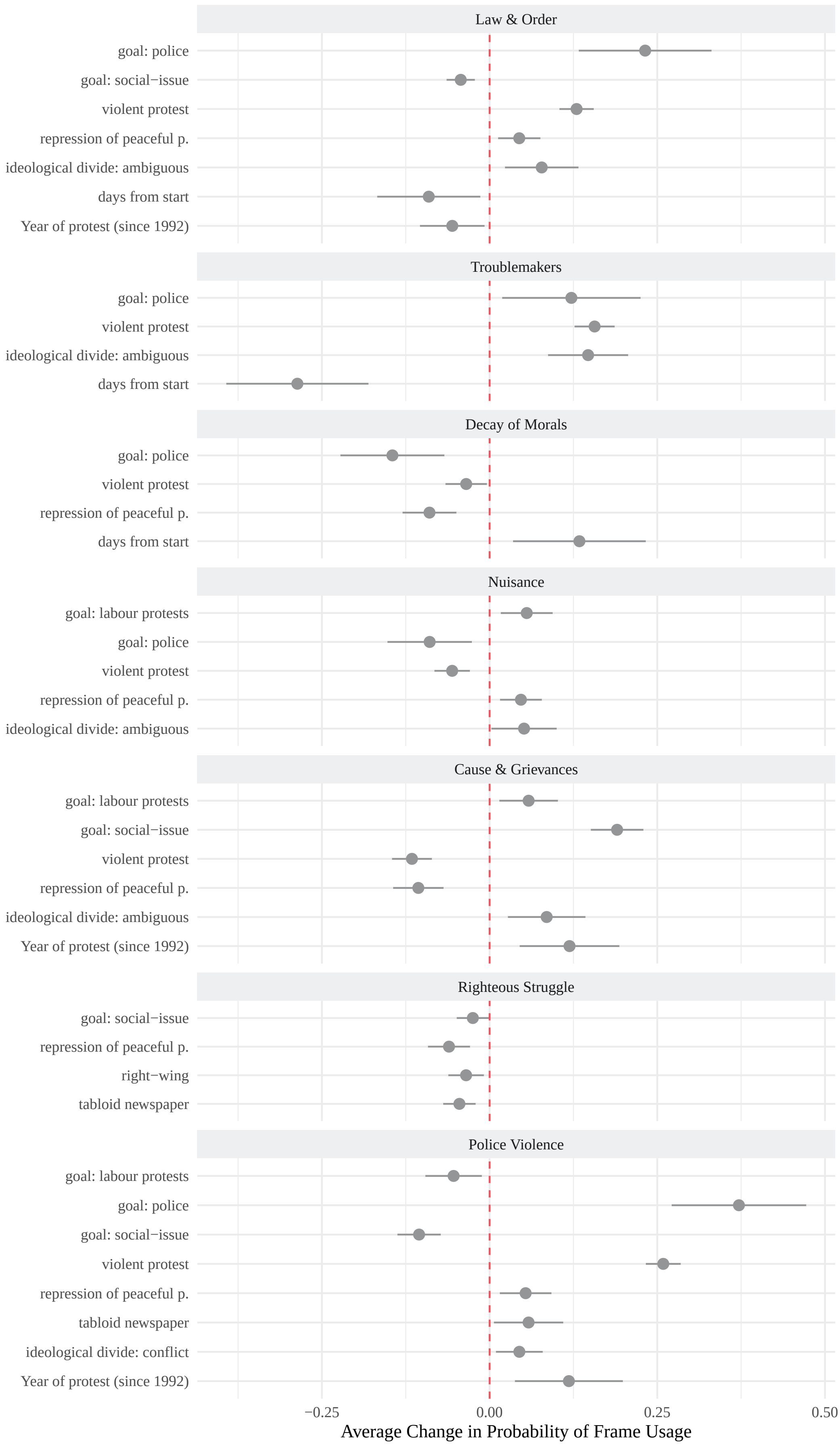


Figure 7.5: Average Marginal Effects for Significant Variables in Individual Frame Models

In the model for aggregated legitimising frames, we saw that *social issue* and *labour* protests are given precedence over the other goals. On the level of the individual legitimising frames, only *social issue* protests are significantly different from the reference category *political* protests in one model. While the frame *cause & grievances* is significantly more likely to be used (AME: 0.19), there is no significant effect in the *righteous struggle* model.

In the model for the aggregated legitimising frame *labour protests* also showed a significant positive effect. Here we see that this effect is solely due to the increased likelihood of *cause & grievances* being used (AME: 0.06). Looking back at the question of a *hierarchy of social struggle*, I concluded that protests against police brutality have a smaller chance of being portrayed in a legitimising frame than *political* protest, while the probability is larger for *anti-war*, *labour* protests and especially *social-issue* protests. This matches what we see in Figure 7.6 for the *cause & grievances* frame. *Righteous struggle*, however, follows a different logic: coverage of *social-issue* and *anti-war* protests appear to use the *righteous struggle* frame less often, only for *labour* is it used more often than for *political* protest. However, the differences are small and none of them is statistically significant.

The remaining frame, *police violence*, which was deemed neither legitimising nor delegitimising, shows the clearest evidence that the different goals play a role when journalists apply one frame or another. Anti-war protests and *political* protests use the frame in the same way. All other demands lead to significantly different usage of this frame. Labour protest and *social-issue* protest are less likely to invoke *police violence* (AME: -0.05, -0.11), whereas anti police brutality protests are, as said, significantly more likely to be covered with this frame (AME: 0.37). Interestingly, this means that the hierarchy of struggles in this model mirrors the one for legitimising frames identified in the last section.

The results for individual frames confirm again the theoretical idea behind *H3a-b*: the goal of a protest matters for how it is covered. Yet, the results are more complicated than one might expect. It is not the case that protest events with some types of goals are covered with frames that are "better" for protesters than other types. Rather, different protest demands seem to lead to a different focus in reporting. Figure 7.6 shows again the distribution of data for the different frames and goals. It shows that the differences between goals are most stark in the *police violence* model. For the other frames, distributions look mostly similar, with only one or two demands following a different trajectory.

Tactics. In terms of protest tactics, the models for individual frames show that the positive effect for delegitimising framing was driven by the usage of *law & order* (AME: 0.13) and *troublemakers* (AME: 0.16). The other two delegitimising frames even have significant negative effects, yet at a smaller magnitude. This makes sense as *law & order* and *troublemakers* both have a focus on violence. *Decay of morals* and *nuisance*, however, are used respectively 3% and 6% less likely once there is violence. In these cases, journalists who want to focus on the downside of an event probably prefer to choose one of the frames focusing directly on violence. This is also shown by the model for *police violence*, which shows that journalists use a healthy amount of scepticism when violence occurs during a protest. AMEs between models are not directly comparable but we can see that the average change of probability, 26 percentage points, is greater than in the models for both *law & order* and *troublemakers*. The model for the usage of *Police Violence* is, therefore, more similar to the ones for delegitimising frames in this regard.

Furthermore, Table 7.6 shows that the negative effect for violence seen in the legitimising model was entirely driven by *cause & grievances*. As violence occurs, journalists are less likely to follow the reasoning of protesters, a finding which basically supports what is already common knowledge. A little surprisingly, the model for *righteous struggle* is the only one to show no significant difference when there is violence during an event. This suggests that journalists portray a cause as just independently from the occurrence of violence — which is somewhat hard to believe. But remember that the

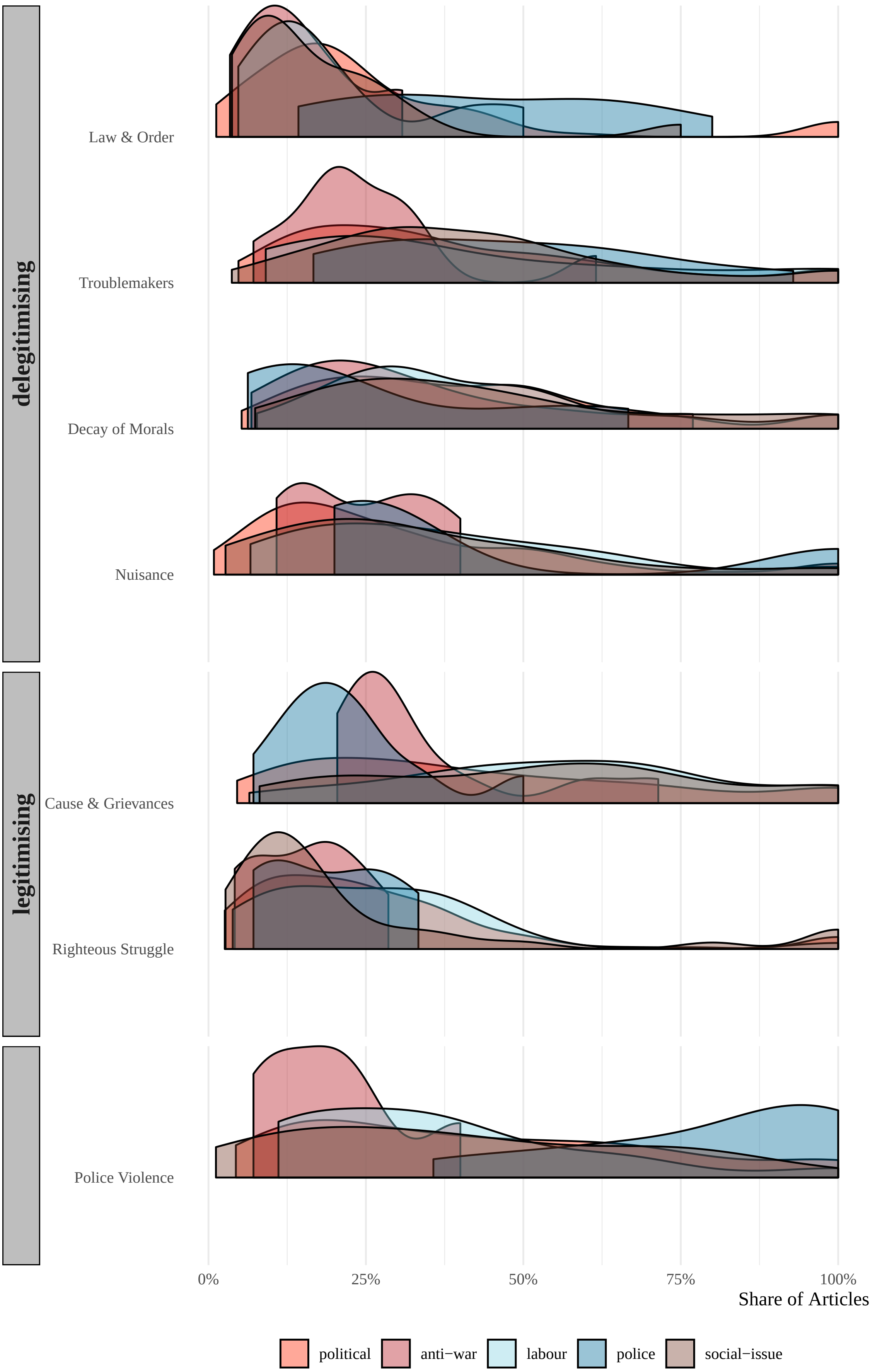


Figure 7.6: Presence of Individual Frames in Percent of Articles About Protests with Different Goals

occurrence of this frame is also independent of the occurrence of other frames. In other words, even when journalists admit that a cause is just, the occurrence of violence still leads them to focus on tactics instead of the message of an event.

State Response. The models for individual frames paint a confusing picture when it comes to the variable *state repression of peaceful protest*. Even though the model for delegitimising frames showed no effect, three of the four frames in the category have significant values here. Yet, probabilities for *law & order* and *nuisance*, for which the estimates are positive, are apparently cancelled out by the negative effect for *decay of morals*. That means while it is 4% more likely that *law & order* is used and 5% more likely that *nuisance* is used, overall delegitimising framing remains on the same level when protesters are arrested as *decay of morals* is used substantially less often (AME: 0.03). It seems that trouble is highlighted when there are arrests, but as the model for *troublemakers* shows no significant effect, protesters are not directly made responsible for it. It is noteworthy that there is a negative effect for *state repression of peaceful protest* in the *decay of morals* model. Remember that this frame highlights the risk that protesters set a bad example for others and disturb the general political consensus. One possible explanation could again be that journalists prefer to employ a frame that highlights trouble when protesters were arrested and feel that this already covers what they would otherwise say about morals. However, the fact that there is also a significant positive effect on the usage of the *nuisance* frame contradicts this idea.

In the two models for legitimising frames, the story from the aggregated model is repeated: there is a significant negative effect on the probabilities that coverage will use *cause & grievances* or *righteous struggle*. Theoretical expectations towards what happens when there is *state repression of peaceful protest* fit even less well for the models of individual frames. The rule seems to be that getting arrested, even when demonstrating peacefully, leads to diminished probabilities of coverage that would be useful for a protest. The only exception from this rule is the *decay of morals* frame, which is used less often in these cases.

7.3.2 Outlet-Level Factors

Newspaper Ideology. On the outlet level, one of the main questions was if a *newspaper's ideology* or the distance between the newspaper's and protesters' ideology, i.e., the *ideological divide*, matters for reporting — with the surprising result that neither of them does. In the models for the aggregated frames, there was no significant effect at all for *newspaper ideology*. This is only slightly different for individual frames. There is a significant negative effect in the models for *righteous struggle*: for articles published in right-leaning outlets, the probability that this frame is used is 4 percentage points lower compared to left-leaning outlets. This result is little surprising as *righteous struggle* basically highlights the advantages protest has for an issue. More surprising is the complete lack of other significant effects of this variable in the other models. There is an effect for *law & order*, suggesting that the frame is *less* common in right-leaning outlets. Alas, the p-value is 0.09, which is usually not considered significant. Overall then, it seems that *newspapers' ideologies* play a small role when it comes to the framing of protest events.

As in the aggregated models, there is no support for the two hypotheses regarding the impact of an *ideological divide* between protest and reporting newspaper. Although all of the models for delegitimising frames show p-values just over the significance threshold of $p < 0.05$. And defying theoretical expectations again, the *police violence* frame, which highlights the responsibility of the police for clashes, is 4% more likely when there is an ideological divide. Yet, the meaning of this frame is not as clear cut as for the others frames and although it promotes a critical view of the police, it also focuses on risks and trouble occurring during a protest. The impression from the last section that the impact of newspaper ideology on protest coverage is incidental instead of systematic is further reinforced by these findings.

Type of Newspaper. Only two of the seven models for the usage of individual frames show any significant effect for type of newspaper: *righteous struggle* is used 5% less likely by tabloid newspaper, *police violence* is used 6% more likely. This means that **H3k** was accepted because *righteous struggle* is used more likely in quality media, while the newspaper type has no significant effect on the other legitimising frame. All delegitimising frames are used regardless of the type of newspaper, which confirms the finding from the last section.

7.3.3 Time-Bound Factors

In terms of time factors, I found for the aggregated frames that the *days since the start* of a protest matter only for delegitimising frames. This story is repeated in the models for individual frames. This provides more evidence that the theory of a *news framing cycle* (Gottlieb, 2015), does not hold up for legitimising framing in the UK: instead of becoming more common after the first phase of a protest, *righteous struggle* is less likely to be used over time, while there is no effect for *cause & grievances*.

For two of the four delegitimising frames, however, the theoretical expectation is further supported: they are less likely to be used the longer a protest is in the past. However, there is no effect for *nuisance* and a positive effect for *decay of morals*. This fits Gottlieb (2015)'s argument who suggests not that specifically delegitimising frames are used in the early stages of protest coverage, but that "conflict frames" (p. 7) are employed most often. When conflict itself becomes less interesting, it appears that *decay of morals*, which highlights that protesters are a fringe group and set a bad example, gains in importance.

The models for individual frames also provide some more evidence that legitimising frames become more important over the years. Compared to 1992, *cause & grievances* is used 12% more likely in 2017. The only delegitimising frame that is affected by a change over the years is *law & order*, which is less likely to be used as time goes on

(AME: -0.06). *Police Violence*, once again, follows the trend of the legitimising frames, as Table 7.6 show a significant positive effect for the *year* variable (AME: 0.12).

7.4 Discussion and Summary

The multilevel logistic regression models presented in this chapter provide further evidence that the importance of the *protest paradigm* for news framing of protest is more conditional than suggested several decades ago (e.g., Kilgo and Harlow, 2019; Lee, 2014; Shahin et al., 2016). Specifically, the systematic analysis of a set of key factors that are expected to shape frame usage in the 5,639 articles in this study showed that the tactics and goals, arrests of protesters, the days between a protest and a report about it play important roles for how protests are framed in the media. Additionally, as time goes on, legitimising framing becomes more likely, as has been shown in the previous chapter. The analysis presented in this chapter is, therefore, an important step toward understanding how different protests are treated by the media.

So which factors do explain the usage of different frames? As it turns out, the answer depends on which frame one looks at. For the aggregated frames, it seems that delegitimising frames were used more arbitrarily. No matter what protesters' goals are, they can expect the same level of delegitimising coverage. The same is true for the arrest of peaceful protesters as well as the outlet and year in which reports were published. Only two factors were confirmed to have any statistically significant influence on the probability with which delegitimising framing was used: violence and how many days after the start of an event an article was published. More violence leads, as expected, to a higher probability of delegitimising framing. Delegitimising coverage is also most likely shortly after a protest event started but becomes less likely as the days pass. Overall, this suggests that the assumption that the *protest paradigm* constitutes a default mindset when it comes to reporting is thus confirmed: except for trying to avoid violence protesters can apparently do little to prevent a delegitimising framing of their efforts.

However, focusing only on delegitimising coverage — as most studies have done so far — only provides part of the picture. A protest report about an event that fights for a *labour* or *social* issue, which remained peaceful, during which no arrests are made, which was published in a broadsheet outlet and which was published in recent years has the highest chance of containing one of the legitimising frames. This leads to one of the most interesting findings of this thesis: the influences that shape the use of more or less legitimising and delegitimising framing in the coverage of protest events are not diametrically opposed — except for violence. Like in the previous chapter, we see that while delegitimising coverage remains overall robust, legitimising coverage is conditioned by factors found in the theory.

Within these broader theory-based categories though, the analysis presented here adds a more nuanced understanding. This is especially noticeable when comparing *decay of morals* frame with the other delegitimising frames: For protests against police brutality, it is used less likely, while all other delegitimising frames become more prominent. Similarly, this frame has a lower probability to be used when violence occurs and when protesters are arrested. It also becomes more likely that the *decay of morals* frame is used the further an event is in the past. Since *decay of morals* is so prominent, these effects cancel out the increased probability that any of the other delegitimising frames is used on the aggregated level.

The picture that delegitimising frames are used more arbitrarily is therefore only partly true. On the aggregate level, it might appear as if there is no difference between *political* protest and protests with other goals. Yet, *Labour* issues are apparently seen more as a nuisance for the press, while reports about anti-police brutality protests focus especially on trouble, but neglect morals or nuisance aspects. Likewise, journalists focus more on trouble and the nuisance aspect when there are arrests but less on *decay of morals*. The increased usage of one of the delegitimising frames is therefore annihilated by the decreased use of another.

For the press, it seems that the frames serve different purposes. When protests are peaceful, they are framed as a *nuisance* or protesters are portrayed as a fringe group with immoral ideas by the *decay of morals* frame. When protesters become violent, they are portrayed as troublemakers who have to be stopped through law & order tactics. Or, in fact, the *police violence* frame is used. Both *troublemakers* and *law & order* are thus most likely to be used for violent protests against police brutality or when arrests occurred, in reports published shortly after an event. Additionally, the *law & order* frame is less likely to be used for *social-issue* protests and is the only delegitimising frame that was used more likely in the 1990s than in the 2010s, once all other factors are taken into account.

A similar pattern emerges for the two legitimising frames: The *cause & grievances* frame was used significantly more likely for *labour* and *social-issue* protests than for protests with other demands, while the *righteous struggle* frame has no significant relationship to any of the goals. *Cause & grievances* is less likely to be used when violence occurs or when protesters are arrested. As found in the last chapter, *cause & grievances* also becomes more prevalent over time, which is not true though for the *righteous struggle* frame: in the model here, which includes all the other mentioned factors, the influence of time on this frame disappears.

Finally, the frame *police violence* is most likely to be used for protests against police violence, if violence occurs or arrests are made during an event. Surprisingly, it is also more likely to be used by tabloid newspapers and when there is an ideological divide between a protest and the reporting newspaper outlet. Again, while this frame is critical of the police, it also highlights the trouble ensuing from a protest. Most often, this means police is made responsible for clashes as protesters are repressed unprovoked. However, tabloids and right-wing outlets also often used this frame to make police responsible for clashes by not being heavy-handed enough. Other factors included, it also becomes more likely over time that protest reports use the *police violence* frame. What is interesting about the police violence frame is that it comprises patterns of protest coverage which have only gained attention in more recent studies (Gerbaudo, 2012; Wasow, 2020) or outside the context of Western democracies (Harlow and Johnson, 2011; Oz, 2016).

How do the different theories perform in the face of the systematic analysis presented here? The most robust finding is that violence predicts the usage of different frames in reports on protest. As violence occurs, journalists are less likely to pay attention to the causes of the protests and more likely to highlight their methods. A fact on which almost all previous studies agree as well. As said, the *protest paradigm* also predicts the kind of framing reasonably well — at least on the aggregated level for delegitimising framing. What the idea of a default mindset when it comes to reporting about protests did not cover though was the growing prevalence of legitimising frames, which co-exist with delegitimising, marginalising and demonising coverage. While other scholars have found the *protest paradigm* to be more conditional though, the analysis here provides that mostly only legitimising framing is influenced by the tested factors.

For the theory of a *hierarchy of social struggle* presented by Kilgo and Harlow (2019), the picture is more mixed. The demands of a protest *do* matter; however, a clear hierarchy did not emerge. The most plausible hierarchy can be observed at the level of aggregated legitimacy frames: social-issue protests are given precedence over protests with a *labour* issue, *anti-war*, *political* protest and finally protests against police brutality. Notably, this order is also different from what previous studies found and what had been theoretically expected (Boyle et al., 2004, 2005, 2012). Furthermore, some of these differences, are not statistically significant. The results for individual frames suggest that rather than following one clear hierarchy, different protest demands seem to lead to a different focus in reporting. Generally though, except for the differences between goals observed in the application of the *police violence* frame, the usage of frames between demands appears rather similar, with only one or two demands following a different trajectory in each frame. The idea that some protests are treated more favourably by the media is, therefore, rejected for the UK case.

The theory presented by Wasow (2020) that *state repression of peaceful protest* leads to more legitimising coverage of said protest could also not be confirmed. In part, this was probably due to a lack of available cases, as not a single case of police violence against peaceful protesters was present for the UK in the MMP data. The employed proxy for state repression, police arrests, causes an effect in the opposite direction: legitimising framing was less likely to be used. Getting arrested, even while peaceful, diminishes the chances for protesters to get their message on the media agenda. The *police violence* frame is also more likely to be used in these cases; however, it is not at all clear if media reports using this frame aide protesters' causes.

The idea that the *organisation* level, which comprises the larger context of the routinised activities such as the goals, policies and structure of a news organisation (Shoemaker and Reese, 2014) influences the framing of protest events was also mostly rejected by the analysis here. The *ideology* of a news outlet matters only for the *righteous struggle* frame, while an ideological divide between protesters and outlet matters only for the *police violence* frame (AME = 0.04). Overall then, it seems that *newspaper ideology* plays a small role when it comes to framing of protest events in the UK.

The difference between tabloid and broadsheet news outlets is surprisingly also rather small according to the analysis presented here. Only two frames are used differently between the newspaper types: the *police violence* frame, which attributes responsibility for clashes during protests to police is more common in tabloid newspapers; the legitimising *righteous struggle* frame, which highlights the positive effect of protests for their respective issues, is more common in broadsheets news. The result is surprising as previous studies assumed that the stark divide between newspaper titles with different ideologies and of different types cater to different audiences and would hence prefer different frames (Cammaerts, 2013; Gavin, 2010). While this might certainly be true for individual protests, and has been found to be the case occasionally during manual coding in this study, the systematic analysis of all protests events and coverage in the UK did not replicate these findings.

Finally, I tested the idea of a *news framing cycle*: just as posed by Gottlieb (2015), coverage of protest begins with a focus on conflict — especially using the *troublemakers* and *law & order* frames. Unlike what was found by Gottlieb (2015), though, these frames are not replaced with legitimising ones later on. In fact, the *righteous struggle* frame is also used less frequently after some time following a protest. Rather, it appears that the conflict frames are replaced by the *decay of morals* frame as time passes after the start of an event. The different finding here might be due to the main force Gottlieb (2015) expects to drive the *news framing cycle*: growing participation. Yet, increased participation after the first day of an event is only common for few high-profile events that stretch over days and months — like the Occupy Wall Street movement for which Gottlieb (2015) developed the theory. Nevertheless, a variable that includes days after a protest started seems to be a worthwhile inclusion for studies that test factors of influence for protest reporting.

A limitation to findings presented in this chapter is that endogeneity effects cannot be ruled out entirely. As discussed in Section 4.3.1 the choice to use data from *Mass Mobilization Project* (Clark and Regan, 2019) for the analysis — which was the only available option — might cause a major endogeneity problem: as independent variables are based on media reports, just as the dependent variable, I cannot rule out that effects presented here are overestimated. Specifically, when, for example, the *troublemakers* frame was used in a report, this might have led Clark and Regan (2019) to code a protest as violent — independently from the actual events. Studies with media-independent data have shown that outright misrepresentation of basic information is relatively rare (McCarthy et al., 1999), however, which means that violence might be overstated but not made up. Yet, this is no guarantee for the validity of the results, especially for the event level variables. As a robustness check, I removed *The Times*, which is the only outlet on which both dependent and independent variables can be based, from the data and calculated the models again. The results presented in Appendix E show that this did not cause any substantial changes.

Chapter 8

Conclusion

Public protest can be a powerful resource for ordinary citizens to participate in the political discourse beyond periodically casting a vote between political actors. In this capacity, protests are crucial to democracy as nearly immediate feedback for representatives in power and as a provider of new issues for the public and political agendas. Provided protesters' messages are heard and understood.

This thesis provided new knowledge on the role of news media in the process of public understanding of protest messages. It did so by analysing the framing of a large-scale news media corpus comprising all reports on domestic protest in the eight most important national newspapers in the United Kingdom between 1992 and 2017. Providing evidence on systematic patterns of coverage and their conditionality on time, event and outlet level factors, the thesis refined and systematised previous findings about how news media reflect the *intended meaning of dissent* (e.g., Kilgo and Harlow, 2019; Lee, 2014; McLeod and Hertog, 1999; Shahin et al., 2016). Overall, the insights presented in this thesis provide a more nuanced and systematic understanding of news media behaviour towards dissenting voices of ordinary citizens. This final chapter of the thesis summarises the key findings, contributions, and acknowledges the study's limitations and outlines avenues for further research.

8.1 Findings

The first important finding is that most news media coverage of protests delegitimises protesters' messages by highlighting the nuisance and trouble caused by the events. This finding places the study in a long tradition of research, commonly subsumed under the term *protest paradigm*. Based on theoretical knowledge about common patterns in newsmaking, the core assumption behind the *protest paradigm* is that journalists tasked with reporting about a protest will first and foremost describe the conflict and controversy accompanying protest events, as these are considered most *newsworthy* (e.g., Shoemaker and Reese, 2014). This thesis showed that delegitimising framing was, and still is, the default when it comes to reporting.

However, that said, legitimising coverage, meaning articles highlighting the messages of protests and the positive influence demonstrations have, has become increasingly common over time. Also, while delegitimising framing is more frequent, it was never entirely dominant. Often, and increasingly so, reports feature frames that highlight protest messages, either individually or in the same report with frames that focus on the method of protests. If the trend seen in the scrutinised 26 years continues, legitimising frames will start to be in the majority in only a few years. And I would argue it is likely that this will be the case. Recent decades have seen changes in the media landscape that explain why legitimising frames are used more often, and these changes are still ongoing. Specifically, protest has been largely normalised to a growing segment of the population and hence audiences. Reporters are in fiercer competition for attention than ever before, meaning that omitting information about an event carries a higher risk of losing audiences than ever before. At the same time, audience demands are also more palpable due to user metrics. All the while, tools for communication become more affordable and protesters become savvier in their usage, making it easier to communicate protest messages to a large audience and making it easier for journalists to gather the information and, again due to competition, harder to ignore.

Furthermore, using the insights into framing effects discussed in this thesis allows for a better understanding of the situation of news framing of protest events in the UK: since between roughly 30% in 1992 and 50% of articles about protest in 2017 use legitimising framing, these frames are likely both accessible and applicable for the audience. In situations like this, research suggests that the competition between several frames will lead individuals to deliberate and form their opinions about an issue based on personal values (e.g., Chong and Druckman, 2007b,c). This contrasts with previous findings that found stronger framing effects when frames about new topics went unchallenged. Consequently, earlier studies concluded that coverage following the *protest paradigm* would lead to delegitimisation, marginalisation and demonisation of protests (e.g., McLeod and Hertog, 1999). Despite the continued pre-eminence of delegitimising coverage, a key result of this thesis is that the 30% to 50% of reporting using alternative narratives ensure that strong framing effects are rather unlikely in the UK. Therefore, the assumption that protest is systematically rendered impotent by the news media is rejected.

The second important finding is that seven frames within and beyond the broad categories of legitimising and delegitimising frames make up the discussion about protest and highlight different aspects of events. From most to least prevalent, these are: *Decay of morals*, that portrays protesters as a fringe group who question legitimate authority and disturb the general political consensus with their actions — which is seen as inherently immoral and threatening to the social order. This frame was present in over a third of all newspaper articles in the sample. The *cause & grievances* frame is the second most salient frame overall and the most salient one in recent years. It takes the substantive content and background of a protest seriously and blames officials for causing or neglecting the issues highlighted by a protest. The *troublemakers* frame highlights the violence and crime during protest events and attributes the risks to public safety and property destruction to protesters. Less salient than the two frames before, it is still present in around a fourth of all protest reports in the dataset.

The *police violence* frame could not be placed into either the legitimising or delegitimising category. Like the *troublemakers* frame, it highlights the violent clashes between protesters and police as well as the ensuing danger to public safety this causes. However, this frame places the responsibility for the chaos on the police. The *nuisance* frame portrays protests as just that: an unnecessary nuisance. Unlike predicted by Di Cicco (2010), this frame does not gain in salience over time and is present in less than a fifth of protest reporting in the UK. The *righteous struggle* frame was the second one that is seen as legitimising protesters' messages. It portrays protest in the most positive way and highlights the benefits of protesters' actions for a cause deemed essential and worthwhile the struggle. The least salient frame, *law & order*, is used in less than 10% of the articles in the data almost throughout the entire time frame. It can be described as the opposite of the *police violence* frame as it praises police for reinstating public order, interestingly without even highlighting the clashes that disturbed it. These are the seven main frames that comprise the coverage of protest events in the UK, which answers the first research question which guided this thesis: *how do British newspapers frame the coverage of domestic protest events?*

To answer RQ2 (*How — if at all — did the framing of protest reporting change over the last 26 years?*), this thesis showed that the frames are salient to a different degree over time and depending on events. This is particularly evident in the salience of the frame *police violence*, the use of which exploded in 2009 when police killed a man during G-20 summit protests. From then on, the press scrutinised police conduct more frequently for several years before the salience returned to pre-2009 levels. Interestingly, the salience of the *law & order* frame increased and fell along with this, but to a far lower degree. A general trend over the years, however, was only observed in the usage of *cause & grievances* and *righteous struggle*, which became significantly more prominent during the period under consideration.

The third important finding is that while few factors condition delegitimising coverage, the degree of legitimising framing is influenced by several factors. The only tested factor which influenced both frame categories was violence: when protests turn violent, legitimising frames are less and delegitimising frames more likely to be used. Additionally, reports published weeks or months after an event are less likely to feature a delegitimising frame. Apart from that, delegitimising framing is applied apparently indiscriminately to protests with different goals, whether arrests are made, by right- and left-wing outlets, by tabloid and broadsheet newspapers and over the years.

Again, this finding supports the idea that the *protest paradigm* constitutes the default mindset of reporters tasked to cover a protest event. But, again, looking at the legitimising frames shows that journalists often depart from the default. Specifically, they do so more likely when covering *labour* and *social-issue* protests, when there is no violence during an event, when no protesters are arrested, when the report is to be published in a broadsheet outlet, and when protesters' ideology is neither left nor right. Among these, the goal of a protest is the strongest predictor: protests that try to highlight a social issue are 16% more likely to be framed legitimising than political protests. Followed by arrests and violence, which both decrease the probability that a legitimising frame is applied.

The final important finding is that the missing conditionality of delegitimising framing can be explained when zooming into the delegitimising category. For the press, it seems that the frames serve different purposes: coverage of protests against police brutality use the *troublemakers* and *law & order* frames more often, but *decay of morals* and *nuisance* are less likely to be used to a similar degree, effectively nullifying the effect on the aggregate level. The likelihood that the *troublemakers* and *law & order* frames are employed are also far higher when violence occurs than it appears on the aggregate level since the likelihood that *decay of morals* or *nuisance* are used are substantially decreased. When arrests are made, journalists use *decay of morals* less and *law & order* as well as *nuisance* more likely. As more time passes after the start of an event, reports feature *troublemakers* and *law & order* less, but *decay of morals* more likely. In other words, while the level of delegitimising framing does not change overall with the factors under consideration, different frames are used in different circumstances.

The two legitimising frames are not used that differently from each other: the conditions of the *cause & grievances* frame are almost identical to the legitimising category overall — except for the difference between tabloid and broadsheet, which is not significant. Unlike on the aggregate level, goals, violence and the year in which an article was published do not impact the usage of *righteous struggle* significantly.

These findings answer the third research question of this thesis (*how do British newspapers frame the coverage of domestic protest events?*): I find that different frames are conditioned by event- and outlet-level as well as time-bound variables. Not one of the variables identified by the literature is negligible. However, there is also not a single factor that conditions the usage of every single one of the individual and aggregated frames. The most robust finding is that violence matters for protest reporting. Yet, it does not significantly condition the usage of the *righteous struggle* frame. Despite the choice of the UK as a case, which was made to test theories about the influence of the *organisation* level of reporting (e.g., Shoemaker and Reese, 2014), outlet level variables play a smaller role than expected: neither the differences between right- and left-wing nor the differences between tabloid and broadsheet outlets have a big impact on protest coverage — even though differences between these outlet types are considered particularly sharp in the UK (e.g., Hallin and Mancini, 2004).

8.2 Contributions

This thesis contributes to existing knowledge in four main ways. Firstly, it places the analysis of protest coverage on empirical data that covers a broader scope and longer time frame than any previous research. The thesis thus fills several gaps identified in the literature on news media coverage of protest. Specifically, most research so far employed relatively narrow case studies that focused on a specific, often radical, protest event or a number of similar events over a short period of time. Instead of selecting specific events, I included all protests that were covered by the selected newspapers into this research. This allows the findings presented here to be generalisable to a greater degree than previous research.

Analysing coverage more systematically also contradicted previous findings, like the insight that the divide between UK newspapers when it comes to ideology and type causes different protest reporting (Cammaerts, 2013; Gavin, 2010). While this might be true for individual events or protests of some selected groups, this study has shown that this is not a general trend for all protest reporting. The scope of this thesis also added to existing knowledge by providing a systematic longitudinal analysis that confirms some of the trends that had been proposed in the literature (e.g., Cottle, 2008). Finally, as the study also moved the strong focus of previous research from the US to a European country, it adds to the insightful evidence from other recent studies (e.g., Coulter et al., 2016; Kyriakidou and Olivas Osuna, 2017; Wouters, 2015b).

Secondly, the thesis made a theoretical contribution by placing the analysis and interpretation of protest coverage patterns on a foundation that better fits advances in the study of media framing. Previous studies about the *protest paradigm* often delivered contradicting results, which are hard to compare though, as they employ different operationalisations of coverage following the paradigm. Since many earlier studies also do not quantify the coverage that follows the paradigm, it is not possible to assess their conclusions about the dominance and influence of this kind of coverage or compare it to other points in time. This thesis also did not study media effects but provided that quantifying how frequently frames are used has a major advantage: since we know from framing effects research that a situation of frame competition leads to deliberation, meaning audience members compare different messages and chooses the most convincing one, knowing how often competing frames are used was key. Specifically, I was able to conclude that despite the pre-eminence of delegitimising coverage, this coverage alone will not lead readers to vilify protesters or render protesters' efforts impotent in the UK.

Thirdly, the thesis contributed a new method to both *identify* and *code* frames in a large corpus. In this thesis, I have developed and successfully implemented a new approach to framing analysis that is both more reliable and valid than previous procedures. Additionally, the approach makes it possible to *code* frames in a large corpus with relatively small demands for manual coding. This methodological contribution has implications beyond the research of protest coverage and even beyond media research as it is, theoretically, capable of handling all text corpora. Specifically, I used the work by Matthes and Kohring (2008), who suggest to code frame elements instead of holistic frames and improved it by using a different dimension reduction technique: factor analysis. I argued that this technique is theoretically better suited than the commonly employed cluster analysis, since it does not force mutually exclusive frames but allows several frames to be present in the same text. I also showed empirically that factor analysis identifies more nuanced and more sensible frames. Additionally, I used the results from factor analysis to *code* frames, which neither Matthes and Kohring (2008) nor others who used the idea of clustering frame elements have done (e.g., David et al., 2011). Through NLP methods and machine learning, I was able to *code* frames in the full corpus of 27,496 newspaper articles instead of using sampling.

Finally, using the outcome of the framing analysis, it was possible to make another empirical contribution. Like a few recent studies have done (e.g., Boyle et al., 2012; Kilgo and Harlow, 2019; Lee, 2014), I tested which factors condition the use of different frames. To the best of my knowledge, the list of factors I tested was the most comprehensive to date, as I added several new ones compared to previous research. Besides the findings above, I was able to speak to several theories about protest coverage: I found some support for the idea of a *protest paradigm*, for the theory of a *hierarchy of social struggle*, which postulates that some protest goals are given precedence over others, and the idea of a *news framing cycle* for protests, which predicts that the time between the start of an event and the publication of a report about it conditions the content of coverage. None of these partly conflicting theories explains the entire picture though. Additionally, theories about the influence of the *organisation* level of reporting and the hypothesis that *state repression of peaceful protest* leads to more legitimising framing were rejected.

Another contribution of this thesis was to highlight procedures and pitfalls of building and cleaning a large-scale database of news media coverage about a specific topic. The ambition was to create a database that is both complete, in the sense that all reports about protest are included, while being free from irrelevant reports. Chapter 5 explained how the reports were first selected through a set of keywords, which underwent a thorough selection process through a pilot study, before being assessed for relevance. Since the main keywords "protest" and "demonstration" have a myriad of different meanings in English, 95% of downloaded data was ultimately discarded as it turned out to be irrelevant to the topic of this thesis. My hope is that the experiences I made with different cleaning procedures might guide others who face similar challenges.

8.3 Limitations and Future Avenues

This dissertation, like most research, also has several potential limitations and weaknesses. First, the nature of the data employed in Section 7 causes concerns for the validity of some of the analysis. As discussed in Section 4.3.1, the choice to use data from the *Mass Mobilization Project* (Clark and Regan, 2019) for the analysis — which was the only available option — might cause a major endogeneity problem: as independent variables are based on media reports, just as the dependent variable, I cannot entirely rule out that the effects presented in Chapter 7 are overestimated. Specifically, when, for example, the *troublemakers* frame was used in a report, this might have led Clark and Regan (2019) to code a protest as violent — independently from the actual events. Studies with media-independent data have shown that outright misrepresentation of basic information is relatively rare (McCarthy et al., 1999), which means that violence might be overstated but not made up. Yet, this is no guarantee for the validity of the results, especially for the event level variables. I presented some robustness checks and mitigation strategies for this problem. But ultimately, the only strategy which would have eliminated the problem would have been to collect media-independent data on the protests — which was deemed to be impossible in the context of a thesis, maybe even at all. This problem is not unique to this thesis, however, and also exists in studies that were published in respected peer-reviewed journals (e.g., Kilgo and Harlow, 2019; Lee, 2014; Shahin et al., 2016).

Second, this research takes into account protest that is reported. As several authors have remarked, this creates a considerable bias as the majority of protest events is generally ignored by the media (e.g., Earl et al., 2004; Oliver and Meyer, 1999; Wouters, 2015b). This thesis is likely to reproduce the media's selection bias of protest described in Section 2.1, in the analyses presented in Chapter 6 and in Chapter 7. In Chapter 7, the problem is even exacerbated as only reports were analysed that could be matched with one of the events in the *Mass Mobilization Project* data (Clark and Regan, 2019). Since this database only includes protests with at least 50 observed participants, the bias against smaller protest is further amplified. This also means that while this thesis makes contributions to the literature about the description of protest in the media, the question of how protests make it into the news was not captured.

Thirdly, this research focuses only on newspapers since data was available throughout the selected time period and the press was, and still is, considered to be an important force in shaping media debates. This might not be considered ideal, though, as it is mostly assumed that newspapers are on the decline not just in terms of circulations but also in terms of influence on the general audience. Getting one's daily news from the internet or broadcasts has become far more common. Including other types of media would have been desirable, alas was not feasible in the context of this thesis.

Finally, while the results of this thesis are generalisable on the UK level, the case study design does not permit speaking about general media trends in other countries.

Several studies in recent times have provided that different reporting styles influence the applicability of theories like the *protest paradigm* in different countries (Dardis, 2006b; Shahin et al., 2016; Wouters, 2015b). Since the main focus of this thesis was to capture change over time and between different events, the trade-off was made in favour of a single country study design.

Future research could address these limitations by conducting a study spanning several countries and including different types of media. This would make it possible to study the effects of cultural differences in reporting and protesting. It would be especially interesting to add inter-media agenda-setting and cross country/media framing analysis to existing knowledge. Where do frames originate and who replicates them? In this thesis, I speculated that the trend of growing importance of legitimising framing is partly due to the influence of social media accounts from protesters who make it easier to gather information about their message. But does the media actually use this information or does the legitimising framing originate elsewhere?

To eliminate the problem of endogeneity in the study of influences of event-level factors on media coverage, it would be worthwhile to use data from protest event databases that do not rely on media coverage or at least triangulate information. Several projects have taken on this task, such as the Crowd Counting Consortium (Pressman and Chenoweth) and Count Love Project (Leung and Perkins). Neither project had the necessary data for this study. However, future research could choose a scope that would make it feasible to employ these datasets rather than using purely news media-based data.

Lastly, while the proposed method to framing analysis was successfully tested in this thesis, there are multiple avenues for potentially enhancing it. While factor analysis proved to be a good choice, several other dimension reduction techniques might lead to similar or even better results. For example, latent Dirichlet allocation, latent class analysis or composite variable analysis. Also, the choice to use machine learning to replicate the results from factor analysis was made based on findings by Burscher et al. (2014) and theoretical expectations. However, it would be good to also test if it made more sense to replicate the coding of frame elements through machine learning and using dimension reduction afterwards. Furthermore, I tested the performance of nine different algorithms and selected the best ones to replicate classification from the manually coded set to the rest of the database. However, the field of machine learning on text data is vast and is also quickly advancing in recent years. The nine tested algorithms are thus but a small subset of machine learning approaches. Word embeddings in particular have been proven to outperform other algorithms in many scenarios and have also become more accessible recently. Employing them can make the analysis of media coverage even more effective, thus helping to identify systematic issues in how the media shape democratic discourse. This would help to eventually overcome such problems.

References

Toril Aalberg, Claes de Vreese, and Jesper Strömbäck. Strategy and game framing. In C. H. de Vreese, Frank Esser, and David Nicolas Hopmann, editors, *Comparing political journalism*, pages 33–49. Routledge, London New York, 2016. ISBN 978-1-138-65585-0 978-1-138-65586-7.

David L. Altheide. *An ecology of communication: Cultural formats of control*. Communication and Social Order : an Aldine de Gruyter series of texts and monographs. Aldine de Gruyter, New York, 1995. ISBN 978-0-202-30533-2.

Edwin Amenta, Thomas Alan Elliott, Nicole Shortt, Amber Celina Tierney, Didem Türkoğlu, and Burrel Vann. From bias to coverage: What explains how news organizations treat social movements. *Sociology Compass*, 11(3):e12460, 2017. ISSN 17519020. doi: 10.1111/soc4.12460.

C. W. Anderson. Between creative and quantified audiences: Web metrics and changing patterns of newswork in local US newsrooms. *Journalism: Theory, Practice & Criticism*, 12(5):550–566, 2011a. ISSN 1464-8849. doi: 10.1177/1464884911402451.

C. W. Anderson. Deliberative, Agonistic, and Algorithmic Audiences: Journalism's Vision of its Public in an Age of Audience Transparency. *International Journal of Communication*, 2011b. ISSN 1932-8036. URL http://ijoc.org/index.php/ijoc/article/view/884/537.

K. T. Andrews and N. Caren. Making the News: Movement Organizations, Media Attention, and the Public Agenda. *American Sociological Review*, 75(6):841–866, 2010. ISSN 0003-1224. doi: 10.1177/0003122410386689.

José Andrés Araiza, Heloisa Aruth Sturm, Pinar Istek, and Mary Angela Bock. Hands Up, Dont Shoot, Whose Side Are You On? Journalists Tweeting the Ferguson Protests. *Cultural Studies - Critical Methodologies*, 16(3):305–312, 2016. doi: 10.1177/1532708616634834.

Laura M. Arpan, Kaysee Baker, Youngwon Lee, Taejin Jung, Lori Lorusso, and Jason Smith. News Coverage of Social Protests and the Effects of Photographs and Prior Attitudes. *Mass Communication and Society*, 9(1):1–20, 2006. ISSN 1520-5436. doi: 10.1207/s15327825mcs09011.

J. Barranco and D. Wisler. Validity and Systematicity of Newspaper Data in Event Analysis. *European Sociological Review*, 15(3):301–322, 1999. ISSN 0266-7215. doi: 10.1093/oxfordjournals.esr.a018265.

Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015. doi: 10.18637/jss.v067.i01.

Frank R Baumgartner, Suzanna De Boef, and Amber E Boydstun. *The decline of the death penalty and the discovery of innocence.* Cambridge University Press, Cambridge; New York, 2008. ISBN 978-0-511-37932-1 978-0-511-37664-1 978-0-511-37845-4 978-0-511-79063-8. URL http://proxy.uqtr.ca/login.cgi?action=login&u=uqtr&db=ebsco&ezurl=http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&AN=220410. OCLC: 1104452731.

O. Baysha. On the dichotomy of corporate vs. alternative journalism: OWS as constructed by echo of Moscow. *International Journal of Communication*, 8(1): 2922–2945, 2014. ISSN 1932-8036.

Andrew Bell and Kelvyn Jones. Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data. *Political Science Research and Methods*, 3(1):133–153, January 2015. ISSN 2049-8470, 2049-8489. doi: 10.1017/psrm.2014.7.

Walter Lance Bennett. Toward a Theory of Press-State Relations in the United States. *Journal of Communication*, 40(2):103–127, 1990. ISSN 0021-9916. doi: 10.1111/j.1460-2466.1990.tb02265.x.

Walter Lance Bennett. An introduction to journalism norms and representations of politics. *Political Communication*, 13(4):373–384, 1996. ISSN 1058-4609. doi: 10.1080/10584609.1996.9963126.

Walter Lance Bennett. New Media Power: The Internet and Global Activism. In Nick Couldry and James Curran, editors, *Contesting media power*, Critical media studies, pages 17–38. Rowman & Littlefield, Lanham, Md., 2003. ISBN 0-7425-2385-3.

Walter Lance Bennett. Communicating global activism: strengths and vulnerabilities of networked politics. In Peter Dahlgren and Wim van de Donk, editors, *Cyberprotest*, pages 109–128. Routledge, London [u.a.], 2010. ISBN 0-415-29784-2.

Walter Lance Bennett and Alexandra Segerberg. The Logic of Connective Action. *Information, Communication & Society*, 15(5):739–768, 2012. ISSN 1369-118X. doi: 10.1080/1369118X.2012.670661.

Walter Lance Bennett and Alexandra Segerberg. Communication in Movements. In Donatella Della Porta and Mario Diani, editors, *The Oxford Handbook of Social Movements*, Oxford Handbooks Ser. Oxford University Press, Incorporated, New York, 2015. ISBN 978-0-19-967840-2. doi: 10.1093/oxfordhb/9780199678402.013.39.

Walter Lance Bennett, Regina G. Lawrence, and Steven Livingston. None Dare Call It Torture: Indexing and the Limits of Press Independence in the Abu Ghraib Scandal. *Journal of Communication*, 56(3):467–485, 2006. ISSN 0021-9916. doi: 10.1111/j.1460-2466.2006.00296.x.

Kenneth Benoit. quanteda: Quantitative Analysis of Textual Data, January 2018. URL http://quanteda.io.

Kenneth Benoit and Akitaka Matsuo. *spacyr: Wrapper to the 'spaCy' 'NLP' Library*, 2020. URL https://CRAN.R-project.org/package=spacyr. R package version 1.2.1.

Kenneth Benoit, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, and Akitaka Matsuo. quanteda: An r package for the quantitative analysis of textual data. *Journal of Open Source Software*, 3(30):774, 2018. doi: 10.21105/joss.00774. URL https://quanteda.io.

Kenneth Benoit, Patrick Chester, and Müller Stefan. *quanteda.classifiers: Models for supervised text classification*, 2020a. URL https://github.com/quanteda/quanteda.classifiers. R package version 0.3.

Kenneth Benoit, David Muhr, and Kohei Watanabe. *stopwords: Multilingual Stopword Lists*, 2020b. URL https://CRAN.R-project.org/package=stopwords. R package version 2.0.

Kenneth Benoit, Kohei Watanabe, Haiyan Wang, Stefan Müller, Patrick O. Perry, Johannes Gruber, Benjamin Lauderdale, and William Lowe. *quanteda.textmodels: Scaling Models and Classifiers for Textual Data*, 2020c. URL https://github.com/quanteda/quanteda.textmodels. R package version 0.9.2.

Dan Berkowitz. Non-Routine News and Newswork: Exploring a What-a-Story. *Journal of Communication*, 42(1):82–94, 1992. ISSN 0021-9916. doi: 10.1111/j.1460-2466.1992.tb00770.x.

E. Bird and R.W Dardenne. Myth, Chronicle and Story: Exploring the Narrative Qualities of News. In James W. Carey, editor, *Media, myths, and narratives*, pages 67–86. Sage Publications, Beverly Hills, Calif., 1988. ISBN 978-0-8039-3049-0.

David M. Blei. Probabilistic topic models. *Communications of the ACM*, 55(4):77–84, April 2012. ISSN 0001-0782, 1557-7317. doi: 10.1145/2133806.2133826. URL https://dl.acm.org/doi/10.1145/2133806.2133826.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent Dirichlet Allocation. *J. Mach. Learn. Res.*, 3:993–1022, 2003. ISSN 1532-4435. URL http://dl.acm.org/citation.cfm?id=944919.944937.

Pablo J. Boczkowski. *News at work: Imitation in an age of information abundance.* University Of Chicago Press, Chicago and London, 2010.

Pablo J. Boczkowski and Limor Peer. The Choice Gap: The Divergent Online News Preferences of Journalists and Consumers. *Journal of Communication*, 61 (5):857–876, 2011. ISSN 00219916. doi: 10.1111/j.1460-2466.2011.01582.x.

Porismita Borah. Conceptual Issues in Framing Theory: A Systematic Examination of a Decade's Literature. *Journal of Communication*, 61(2):246–263, 2011. ISSN 00219916. doi: 10.1111/j.1460-2466.2011.01539.x.

Endre Borbáth and Theresa Gessler. Different worlds of contention? Protest in Northwestern, Southern and Eastern Europe. *European Journal of Political Research*, pages 1475–6765.12379, February 2020. ISSN 0304-4130, 1475-6765. doi: 10.1111/1475-6765.12379.

Jelle W. Boumans and Damian Trilling. Taking Stock of the Toolkit. *Digital Journalism*, 4(1):8–23, 2015. ISSN 2167-0811. doi: 10.1080/21670811.2015.1096598.

danah boyd. Social Network Sites as Networked Publics: Affordances, Dynamics, and Implications. In Zizi Papacharissi, editor, *A networked self*, pages 39–58. Routledge, New York, 2011. ISBN 978-0-415-80180-5.

Amber E. Boydstun. *Making the news: Politics, the media, and agenda setting.* Univ. of Chicago Press, Chicago [u.a.], 2013. ISBN 0-226-06560-X.

Amber E. Boydstun, Anne Hardy, and Stefaan Walgrave. Two Faces of Media Attention: Media Storm Versus Non-Storm Coverage. *Political Communication*, 31(4): 509–531, October 2014. ISSN 1058-4609, 1091-7675. doi: 10.1080/10584609.2013.875967. URL http://www.tandfonline.com/doi/abs/10.1080/10584609.2013.875967.

Jules Boykoff. Framing Dissent: Mass-Media Coverage of the Global Justice Movement. *New Political Science*, 28(2):201–228, 2006. ISSN 0739-3148. doi: 10.1080/07393140600679967.

Jules Boykoff. Limiting Dissent: The Mechanisms of State Repression in the USA. *Social Movement Studies*, 6(3):281–310, December 2007. ISSN 1474-2837, 1474-2829. doi: 10.1080/14742830701666988. URL https://www.tandfonline.com/doi/full/10.1080/14742830701666988.

Michael P. Boyle and Cory L. Armstrong. Measuring Level of Deviance: Considering the Distinct Influence of Goals and Tactics on News Treatment of Abortion Protests. *Atlantic Journal of Communication*, 17(4):166–183, 2009. ISSN 1545-6870. doi: 10.1080/15456870903156134.

Michael P. Boyle, Michael R. McCluskey, Narayan Devanathan, Susan E. Stein, and Douglas McLeod. The Influence of Level of Deviance and Protest Type on Coverage of Social Protest in Wisconsin from 1960 to 1999. *Mass Communication and Society*, 7(1):43–60, 2004. ISSN 1520-5436. doi: 10.1207/s15327825mcs07014.

Michael P. Boyle, M. R. McCluskey, Douglas M. McLeod, and S. E. Stein. Newspapers and Protest: An Examination of Protest Coverage from 1960 to 1999. *Journalism & Mass Communication Quarterly*, 82(3):638–653, 2005. ISSN 1077-6990. doi: 10.1177/107769900508200310.

Michael P. Boyle, Douglas M. McLeod, and C. L. Armstrong. Adherence to the Protest Paradigm: The Influence of Protest Goals and Tactics on News Coverage in U.S. and International Newspapers. *The International Journal of Press/Politics*, 17(2): 127–144, 2012. ISSN 1940-1612. doi: 10.1177/1940161211433837.

Monica Brasted. Framing Protest: The Chicago Tribune and the New York Times during the 1968 Democratic Convention. *Atlantic Journal of Communication*, 13(1): 1–25, 2005. ISSN 1545-6870. doi: 10.1207/s15456889ajc13011.

Timothy A. Brown. *Confirmatory factor analysis for applied research*. Methodology in the social sciences. The Guilford Press, New York ; London, second edition edition, 2015. ISBN 978-1-4625-1779-4.

Mark L. Bryan and Stephen P. Jenkins. Multilevel Modelling of Country Effects: A Cautionary Tale. *European Sociological Review*, 32(1):3–22, February 2016. ISSN 0266-7215, 1468-2672. doi: 10.1093/esr/jcv059. URL https://academic.oup.com/esr/article-lookup/doi/10.1093/esr/jcv059.

Björn Burscher, Daan Odijk, Rens Vliegenthart, Maarten de Rijke, and Claes H. de Vreese. Teaching the Computer to Code Frames in News: Comparing Two Supervised Machine Learning Approaches to Frame Analysis. *Communication Methods and Measures*, 8(3):190–206, 2014. ISSN 1931-2458. doi: 10.1080/19312458.2014.937527.

P. Burstein. Why Estimates of the Impact of Public Opinion on Public Policy are Too High: Empirical and Theoretical Implications. *Social Forces*, 84(4):2273–2289, 2006. ISSN 00377732, 15347605. doi: 10.1353/sof.2006.0083.

Karl-Heinz Bäuml. Inhibitory Processes. In Randolph Menzel and John H. Byrne, editors, *Learning and Memory*, pages 195–220. Elsevier, Oxford, 2008.

Bart Cammaerts. Protest logics and the mediation opportunity structure. *European Journal of Communication*, 27(2):117–134, 2012. ISSN 0267-3231. doi: 10.1177/0267323112441007.

Bart Cammaerts. The Mediation of Insurrectionary Symbolic Damage. *The International Journal of Press/Politics*, 18(4):525–548, 2013. ISSN 1940-1612. doi: 10.1177/1940161213496283.

Bart Cammaerts, Alice Mattoni, and Patrick McCurdy, editors. *Mediation and Protest Movements*. Intellect, 2013.

Joseph N. Cappella and Kathleen Hall Jamieson. *Spiral of cynicism: The press and the public good*. Oxford University Press, New York, 1997. ISBN 0-19-509064-0.

Dallas Card, Amber E. Boydstun, Justin H. Gross, Philip Resnik, and Noah A. Smith. The media frames corpus: Annotations of frames across issues. *Proceedings of Association for Computational Linguistics Conference (ACL)*, 2015.

Kevin M. Carragee and Wim Roefs. The Neglect of Power in Recent Framing Research. *Journal of Communication*, 54(2):214–233, 2004. doi: 10.1111/j.1460-2466.2004.tb02625.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1460-2466.2004.tb02625.x.

Douglass Cater. *The fourth branch of government*. Houghton Mifflin, Boston, 1959.

Andrew Chadwick. *The hybrid media system: politics and power*. Oxford studies in digital politics. Oxford University Press, New York, NY, second edition edition, 2017. ISBN 978-0-19-069673-3 978-0-19-069672-6.

Andrew Chadwick and James Dennis. Social Media, Professional Media and Mobilisation in Contemporary Britain: Explaining the Strengths and Weaknesses of the Citizens' Movement 38 Degrees. *Political Studies*, 65(1):42–60, 2017. ISSN 0032-3217. doi: 10.1177/0032321716631350.

Joseph Man Chan and Chi-Chuan Lee. Journalistic 'Paradigms' of Civil Protests: A Case Study of Hong Kong. In Andrew Arno, editor, *The news media in national and international conflict*, A Westview special study, pages 183–202. Westview Press, Boulder, Colo., 1984. ISBN 0-86531-776-3.

Jonathan Chang, Sean Gerrish, Chong Wang, Jordan Boyd-graber, and David Blei. Reading tea leaves: How humans interpret topic models. In Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta, editors, *Advances in Neural Information Processing Systems*, volume 22. Curran Associates, Inc., 2009. URL https://proceedings.neurips.cc/paper/2009/file/f92586a25bb3145facd64ab20fd554ff-Paper.pdf.

Malika Charrad, Nadia Ghazzali, Véronique Boiteau, and Azam Niknafs. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software*, 61(6):1–36, 2014. URL http://www.jstatsoft.org/v61/i06/.

Dennis Chong and James N. Druckman. Framing Public Opinion in Competitive Democracies. *American Political Science Review*, 101(04), 2007a. ISSN 0003-0554. doi: 10.1017/S0003055407070554.

Dennis Chong and James N. Druckman. Framing Theory. *Annual Review of Political Science*, 10(1):103–126, 2007b. ISSN 1094-2939. doi: 10.1146/annurev.polisci.10.072805.103054.

Dennis Chong and James N. Druckman. A Theory of Framing and Opinion Formation in Competitive Elite Environments. *Journal of Communication*, 57(1):99–118, 2007c. ISSN 00219916. doi: 10.1111/j.1460-2466.2006.00331.x.

Hsiang Iris Chyi and Maxwell McCombs. Media Salience and the Process of Framing: Coverage of the Columbine School Shootings. *Journalism & Mass Communication Quarterly*, 81(1):22–35, March 2004. ISSN 1077-6990, 2161-430X. doi: 10.1177/107769900408100103.

David Clark and Patrick Regan. Mass mobilization protest data, 2019. URL https://massmobilization.github.io/.

Jacob Cohen. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1):37–46, April 1960. ISSN 0013-1644, 1552-3888.

doi: 10.1177/001316446002000104. URL http://journals.sagepub.com/doi/10.1177/001316446002000104.

S. Coleman. Online Campaigning. *Parliamentary Affairs*, 54(4):679–688, 2001. ISSN 0031-2290. doi: 10.1093/parlij/54.4.679.

Timothy E. Cook. *Governing with the news: The news media as a political institution.* Studies in communication, media, and public opinion. University Of Chicago Press, Chicago, second edition edition, 2005. ISBN 0-226-02668-X.

Timothy E. Cook. The News Media as a Political Institution: Looking Backward and Looking Forward. *Political Communication*, 23(2):159–171, 2006. ISSN 1058-4609. doi: 10.1080/10584600600629711.

Danielle Sarver Coombs, Cheryl Ann Lambert, David Cassilo, and Zachary Humphries. Flag on the Play: Colin Kaepernick and the Protest Paradigm. *Howard Journal of Communications*, 31(4):317–336, August 2020. ISSN 1064-6175, 1096-4649. doi: 10.1080/10646175.2019.1567408. URL https://www.tandfonline.com/doi/full/10.1080/10646175.2019.1567408.

C. Corrigall-Brown and R. Wilkes. Picturing Protest: The Visual Framing of Collective Action by First Nations in Canada. *American Behavioral Scientist*, 56(2):223–243, 2012. ISSN 0002-7642. doi: 10.1177/0002764211419357.

Simon Cottle, editor. *Media organisation and production.* Media in focus. Sage, London u.a., 1. publ edition, 2003. ISBN 0-7619-7493-8.

Simon Cottle. Reporting demonstrations: The changing media politics of dissent. *Media, Culture & Society*, 30(6):853–872, 2008. ISSN 0163-4437. doi: 10.1177/0163443708096097.

C. Coulter, H. Browne, R. Flynn, V. Hetherington, and G. Titley. 'These people protesting might not be so strident if their own jobs were on the line: Representations of the 'economic consequences of opposition to the Iraq war in the Irish national press. *Media, War & Conflict*, 9(2):113–136, 2016. ISSN 1750-6352. doi: 10.1177/1750635216637635.

Jasper Cox. New research: Some 198 UK local newspapers have closed since 2005. *PressGazette*, December 2016.

James Curran and Jean Seaton. *Power without responsibility: The press, broadcasting, and new media in Britain*. Routledge, London and New York, repr. 6th ed. edition, 2007. ISBN 0-415-24390-4.

James Curran, Sharon Coen, Toril Aalberg, Kaori Hayashi, Paul K. Jones, Sergio Splendore, Stylianos Papathanassopoulos, David Rowe, and Rod Tiffen. Internet revolution revisited: A comparative study of online news. *Media, Culture & Society*, 35(7):880–897, 2013. ISSN 0163-4437. doi: 10.1177/0163443713499393.

Russell J. Dalton. *Democratic Challenges, Democratic Choices*. Oxford University Press, 2004. ISBN 978-0-19-926843-6. doi: 10.1093/acprof:oso/9780199268436.001.0001.

Paul D'Angelo. News Framing as a Multiparadigmatic Research Program: A Response to Entman. *Journal of Communication*, 52(4):870–888, 2002. ISSN 00219916. doi: 10.1111/j.1460-2466.2002.tb02578.x. URL http://dx.doi.org/10.1111/j.1460-2466.2002.tb02578.x.

Paul D'Angelo and Jim A. Kuypers, editors. *Doing news framing analysis: Empirical and theoretical perspectives*. Communication series. Routledge, New York and London, 2010. ISBN 0-203-86446-8.

Frank E. Dardis. Marginalization Devices in U.S. Press Coverage of Iraq War Protest: A Content Analysis. *Mass Communication and Society*, 9(2):117–135, 2006a. ISSN 1520-5436. doi: 10.1207/s15327825mcs09021.

Frank E. Dardis. Military Accord, Media Discord: A Cross-National Comparison of UK vs US Press Coverage of Iraq War Protest. *International Communication Gazette*, 68(5-6):409–426, 2006b. ISSN 1748-0485. doi: 10.1177/1748048506068719.

Clarissa C. David, Jenna Mae Atun, Erika Fille, and Christopher Monterola. Finding Frames: Comparing Two Methods of Frame Analysis. *Communication Methods and Measures*, 5(4):329–351, October 2011. ISSN 1931-2458, 1931-2466. doi: 10.1080/19312458.2011.624873. URL http://www.tandfonline.com/doi/abs/10.1080/19312458.2011.624873.

Aeron Davis. Public relations, news production and changing patterns of source access in the British national media. *Media, Culture & Society*, 22(1):39–59, 2000. ISSN 0163-4437. doi: 10.1177/016344300022001003.

Aeron Davis. News, Public Relations and Power. In Simon Cottle, editor, *News, public relations and power*, Media in focus, pages 109–128. SAGE Publications Ltd, 2003. ISBN 978-0-7619-7496-3. doi: 10.4135/9781446221594. URL http://sk.sagepub.com/books/news-public-relations-and-power.

Daniel Dayan and Elihu Katz. *Media events: The live broadcasting of history*. Harvard University Press, Cambridge Mass, 1992. ISBN 978-0-674-55956-1.

Arnold S. de Beer. Ecquid Novi – the search for a definition. *Ecquid Novi: African Journalism Studies*, 25(2):186–209, 2010. ISSN 0256-0054. doi: 10.1080/02560054.2004.9653294.

Brooks De Cillia and Patrick McCurdy. No Surrender. No Challenge. No Protest Paradigm: A Content Analysis of the Canadian News Media Coverage of the "Yellow Vest Movement" and the "United We Roll Convoy". *Canadian Review of Sociology/Revue canadienne de sociologie*, 57(4):656–680, November 2020. ISSN 1755-6171, 1755-618X. doi: 10.1111/cars.12304. URL https://onlinelibrary.wiley.com/doi/10.1111/cars.12304.

Claes H De Vreese. News framing: Theory and typology. *Information design journal+ document design*, 13(1):51–62, 2005.

Claes H. de Vreese, Jochen Peter, and Holli A. Semetko. Framing Politics at the Launch of the Euro: A Cross-National Comparative Study of Frames in the News. *Political Communication*, 18(2):107–122, January 2001. ISSN 1058-4609. doi: 10.1080/105846001750322934.

David Deacon. Yesterday's Papers and Today's Technology. *European Journal of Communication*, 22(1):5–25, 2007. ISSN 0267-3231. doi: 10.1177/0267323107073743.

Donatella Della Porta. Repertoires of Contention. In David A. Snow, Donatella Della Porta, Bert Klandermans, and Doug McAdam, editors, *The Wiley-Blackwell Encyclopedia of Social and Political Movements*, pages 43–46. Blackwell Publishing Ltd, Oxford, UK, 2013. ISBN 978-1-4051-9773-1. doi: 10.1002/9780470674871.wbespm178.

Matthew J. Denny and Arthur Spirling. Text Preprocessing For Unsupervised Learning: Why It Matters, When It Misleads, And What To Do About It. *Political Analysis*, 26(2):168–189, April 2018. ISSN 1047-1987, 1476-4989. doi: 10.1017/pan.2017.44.

Benjamin H. Detenber, Melissa R. Gotlieb, Douglas M. McLeod, and Olga Malinkina. Frame Intensity Effects of Television News Stories About a High-Visibility Protest Issue. *Mass Communication and Society*, 10(4):439–460, 2007. ISSN 1520-5436. doi: 10.1080/15205430701580631.

Damon T. Di Cicco. The Public Nuisance Paradigm: Changes in Mass Media Coverage of Political Protest since the 1960s. *Journalism & Mass Communication Quarterly*, 87(1):135–153, 2010. ISSN 1077-6990. doi: 10.1177/107769901008700108.

Murray Dick. Search engine optimisation in UK news production. *Journalism Practice*, 5(4):462–477, 2011. ISSN 1751-2786. doi: 10.1080/17512786.2010.551020.

Paul DiMaggio, Manish Nag, and David Blei. Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. *Poetics*, 41(6):570–606, 2013. doi: 10.1016/j.poetic.2013.08.004.

Anthony Downs. Up and down with ecology: The issue attention cycle. *The Public Interest*, (28):38–51, 1972.

JamesN. Druckman. The Implications of Framing Effects for Citizen Competence. *Political Behavior*, 23(3):225–256, 2001. ISSN 0190-9320. doi: 10.1023/A:1015006907312. URL http://dx.doi.org/10.1023/A%3A1015006907312.

Jennifer Earl, Andrew Martin, John D. McCarthy, and Sarah A. Soule. The Use of Newspaper Data in the Study of Collective Action. *Annual Review of Sociology*, 30 (1):65–80, 2004. ISSN 0360-0572. doi: 10.1146/annurev.soc.30.012703.110603.

L. Edgerly, A. Toft, and M. L. Veden. Social Movements, Political Goals, and the May 1 Marches: Communicating Protest in Polysemous Media Environments. *The International Journal of Press/Politics*, 16(3):314–334, 2011. ISSN 1940-1612. doi: 10.1177/1940161211398480.

Mohamad Hamas Elmasry and Mohammed el Nawawy. Do black lives matter?: A content analysis of new york times and st. louis post-dispatch coverage of michael brown protests. *Journalism Practice*, 11(7):857–875, August 2017. ISSN 1751-2786, 1751-2794. doi: 10.1080/17512786.2016.1208058. URL https://www.tandfonline.com/doi/full/10.1080/17512786.2016.1208058.

Christian Elmelund-Præstekær and Charlotte Wien. What's the Fuss About? The Interplay of Media Hypes and Politics. *The International Journal of Press/Politics*, 13(3):247–266, 2008. ISSN 1940-1612. doi: 10.1177/1940161208319292.

Sally Emerson. Why i've become a placard waver! from post offices to parliament, the middle classes are in revolt. *Daily Mail*, 2008, 2008-03-04.

Kathleen L. Endres. “Help-Wanted Female”: Editor & Publisher Frames a Civil Rights Issue. *Journalism & Mass Communication Quarterly*, 81(1):7–21, 2004. ISSN 1077-6990. doi: 10.1177/107769900408100102.

B. Enjolras, K. Steen-Johnsen, and D. Wollebaek. Social media and mobilization to offline demonstrations: Transcending participatory divides? *New Media & Society*, 15(6):890–908, 2013. ISSN 1461-4448. doi: 10.1177/1461444812462844.

Robert M. Entman. Framing: Toward Clarification of a Fractured Paradigm. *Journal of Communication*, 43(4):51–58, 1993. ISSN 0021-9916. doi: 10.1111/j.1460-2466.1993.tb01304.x. URL https://doi.org/10.1111/j.1460-2466.1993.tb01304.x.

Robert M. Entman. *Projections of Power: Framing News, Public Opinion, and U.S. Foreign Policy*. Studies in Communication, Media, and Pub. University Of Chicago Press, Chicago, 2004. ISBN 978-0-226-21073-5. URL http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=298799.

Robert M. Entman. Framing Bias: Media in the Distribution of Power. *Journal of Communication*, 57(1):163–173, 2007. ISSN 00219916. doi: 10.1111/j.1460-2466.2006.00336.x.

Robert M. Entman. Framing Media Power. In Paul D’Angelo and Jim A. Kuypers, editors, *Doing news framing analysis*, Communication series, pages 331–355. Routledge, New York and London, 2010. ISBN 0-203-86446-8.

Robert M. Entman, Jörg Matthes, and Lynn Pellicano. Nature, Sources, and Effects of News Framing. In Karin Wahl-Jorgensen and Thomas Hanitzsch, editors, *The handbook of journalism studies*, International Communication Association handbook series, pages 163–173. Routledge, New York, 2009. ISBN 1-4106-1806-4.

Brian Everitt. *Cluster analysis*, volume 848 of *Wiley series in probability and statistics*. Wiley, Chichester, West Sussex, U.K., 5. ed. edition, January 2011. ISBN 978-0-470-74991-3.

Jessica T. Feezell. Agenda Setting through Social Media: The Importance of Incidental News Exposure and Social Filtering in the Digital Era. *Political Research Quarterly*, 71(2):482–494, June 2018. ISSN 1065-9129, 1938-274X. doi: 10.1177/1065912917744895. URL http://journals.sagepub.com/doi/10.1177/1065912917744895.

Andy Field, Jeremy Miles, and Zoë Field. *Discovering Statistics Using R*. Sage, Los Angeles, Calif., reprint edition, 2012. ISBN 978-1-4462-0046-9. OCLC: 854989686.

Dana R. Fisher, Kenneth T. Andrews, Neal Caren, Erica Chenoweth, Michael T. Heaney, Tommy Leung, L. Nathan Perkins, and Jeremy Pressman. The science of contemporary street protest: New efforts in the United States. *Science Advances*, 5(10), October 2019. ISSN 2375-2548. doi: 10.1126/sciadv.aaw5461.

Carlos G. Forero, Alberto Maydeu-Olivares, and David Gallardo-Pujol. Factor Analysis with Ordinal Indicators: A Monte Carlo Study Comparing DWLS and ULS Estimation. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(4):625–641, October 2009. ISSN 1070-5511, 1532-8007. doi: 10.1080/10705510903203573. URL http://www.tandfonline.com/doi/abs/10.1080/10705510903203573.

Ron Francisco. Protest and coercion data. 28 european countries from 1980 through 1995, n.d. URL https://ronfran.ku.edu/data/index.html.

Adam Fresco and Richard Ford. Police 'snatch squads' ready to arrest violent demonstrators; home secretary changes tack on water cannon. *The Times*, 2010, 2010-12-14.

Katja Friedrich, Till Keyling, and Hans-Bernd Brosius. Gatekeeping Revisited. In Gerhard Vowe and Philipp Henn, editors, *Political communication in the online world*, Routledge research in political communication, pages 59–72. Routledge, New York and London, 2016. ISBN 978-1-315-70749-5.

Johan Galtung and Mari Holmboe Ruge. The Structure of Foreign News. *Journal of Peace Research*, 2(1):64–91, 1965. ISSN 00223433.

Matthias Gamer, Jim Lemon, and Ian Fellows Puspendra Singh. *irr: Various Coefficients of Interrater Reliability and Agreement*, 2019. URL https://CRAN.R-project.org/package=irr. R package version 0.84.1.

William A. Gamson and Andre Modigliani. Media Discourse and Public Opinion on Nuclear Power: A Constructionist Approach. *American Journal of Sociology*, 95(1): 1–37, 1989. ISSN 0002-9602. doi: 10.1086/229213.

William A. Gamson, David Croteau, William Hoynes, and Theodore Sasson. Media Images and the Social Construction of Reality. *Annual Review of Sociology*, 18: 373–393, 1992. ISSN 03600572, 15452115.

Oscar H. Gandy. *Beyond agenda setting: Information subsidies and public policy*. Communication and information science. Ablex Pub. Co., Norwood, N.J., 1982. ISBN 978-0-89391-096-9.

Herbert J. Gans. *Deciding what's news: A study of CBS evening news, NBC nightly news,'Newsweek' and 'Time'*, volume 12 of *Communication and society*. Constable, London, 25. anniversary ed. edition, 1980. ISBN 0-09-463390-8.

Neil T. Gavin. *Press and television in British politics: Media, money and mediated democracy*. Palgrave Macmillan, Basingstoke [England] and New York, 2007. ISBN 1-349-51010-6.

Neil T. Gavin. Pressure Group Direct Action on Climate Change: The Role of the Media and the Web in Britain-A Case Study. *The British Journal of Politics & International Relations*, 12(3):459–475, 2010. ISSN 13691481. doi: 10.1111/j.1467-856X.2010.00411.x.

Andrew Gelman and Jennifer Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, Leiden, 2006. ISBN 978-0-511-26878-6. URL http://www.SLQ.eblib.com.au/patron/FullRecord.aspx?p=288457. OCLC: 437176913.

Paolo Gerbaudo. *Tweets and the streets: Social media and contemporary activism*. Pluto Press, London, 2012. ISBN 978 0 7453 3249 9. doi: 10.2307/j.ctt183pdzs. URL http://www.jstor.org/stable/10.2307/j.ctt183pdzs.

B. G. Ghobrial and K. G. Wilkins. The politics of political communication: Competing news discourses of the 2011 Egyptian protests. *International Communication Gazette*, 77(2):129–150, 2015. ISSN 1748-0485. doi: 10.1177/1748048514564027.

Teresa Gil-Lopez. Mainstream Protest Reporting in the Contemporary Media Environment: Exploring (In)Stability and Adherence to Protest Paradigm From 1998 to 2017. *Journalism & Mass Communication Quarterly*, page 107769902098478, December 2020. ISSN 1077-6990, 2161-430X. doi: 10.1177/1077699020984783. URL http://journals.sagepub.com/doi/10.1177/1077699020984783.

Fabrizio Gilardi, Charles R. Shipan, and Bruno Wüest. Policy Diffusion: The Issue-Definition Stage. *American Journal of Political Science*, 65(1):21–35, January 2021. ISSN 0092-5853, 1540-5907. doi: 10.1111/ajps.12521. URL https://onlinelibrary.wiley.com/doi/10.1111/ajps.12521.

Todd Gitlin. *The whole world is watching: Mass media in the making & unmaking of the New Left*. University of California Press, Berkeley, 1980. ISBN 978-0-520-04024-3.

Glasgow University Media Group. *War and peace news*. Open University Press, Milton Keynes and Philadelphia, 1985. ISBN 978-0-335-10598-4.

Erving Goffman. *Frame analysis: An essay on the organization of experience*, volume CN 372 of *Harper colophon books*. Harper & Row, New York, 1974. ISBN 0-930350-91-X.

Jill Gordon. John Stuart Mill and the \dqMarketplace of Ideas\dq. *Social Theory and Practice*, 23(2):235–249, 1997. ISSN 0037802X, 2154123X.

Julian Gottlieb. Protest news framing cycle: How the New York times covered occupy wall street. *International Journal of Communication*, 9(1):231–253, 2015. ISSN 1932-8036.

Arthur C. Graesser, Danielle S. McNamara, and Max M. Louwerse. Methods of automated text analysis. In Michael L. Kamil, P. David Pearson, Elizabeth Birr Moje, and Peter P. Afflerbach, editors, *Handbook of Reading Research.*, pages 34–53. Routledge, New York, 2011. ISBN 978-0-8058-5342-1.

Edgar Grande, Swen Hutter, Hanspeter Kriesi, Martin Dolezal, Johan Hellström, and Simon Maag. Poldem-Protest dataset on EU issues, 2020. URL https://poldem.eui.eu/download/protest-events/.

Rob Griffin, Jamie Wilson, and Angelique Chrisafis. Fuel crisis: Cost of dispute could top pounds 1bn, say firms: Protests called off 'in the nick of time' to avoid factory closures and the laying off of workers * tourism and rural schools badly hit. *The Guardian*, 2000, 2000-09-16.

J. Grimmer and B. M. Stewart. Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, 21(3):267–297, 2013. ISSN 1047-1987. doi: 10.1093/pan/mps028.

Johannes Gruber. *LexisNexisTools. An R package for working with newspaper data from 'LexisNexis'*, 2021. URL https://github.com/JBGruber/LexisNexisTools. R package version 0.3.3.

Bettina Grün and Kurt Hornik. topicmodels: An R package for fitting topic models. *Journal of Statistical Software*, 40(13):1–30, 2011. doi: 10.18637/jss.v040.i13.

Jürgen Habermas. *The structural transformation of the public sphere: An inquiry into a category of bourgeois society.* Polity, Cambridge, UK, repr edition, 1989. ISBN 0-7456-0274-6.

Stuart Hall, Chas Critcher, Tony Jefferson, John Clarke, and Brian Roberts. *Policing the crisis: Mugging, the state, and law and order.* Palgrave Macmillan, Basingstoke, Hampshire, 1978. ISBN 978-1-137-00719-3.

Daniel C. Hallin. *The \dquncensored war\dq: The media and Vietnam.* Oxford University Press, New York, 1986a. ISBN 978-0-19-503814-9.

Daniel C. Hallin. *The "Uncensored War": The Media and Vietnam.* Oxford University Press, New York, 1986b. ISBN 978-0-19-503814-9.

Daniel C. Hallin and Paolo Mancini. *Comparing media systems: Three models of media and politics.* Communication, society, and politics. Cambridge University Press, Cambridge and New York, 2004. ISBN 978-0-511-21075-4.

James D. Halloran, Philip Ross Courtney Elliott, and Graham Murdock. *Demonstrations and communication: A case study.* Penguin special. Penguin, Harmondsworth, 1970. ISBN 978-0-14-052282-2.

James Hamilton. *All the news that's fit to sell: How the market transforms information into news.* Princeton University Press, Princeton (N.J.), 2004. ISBN 978-0-691-11680-8.

Raymond A. Harder, Julie Sevenans, and Peter van Aelst. Intermedia Agenda Setting in the Social Media Age: How Traditional Players Dominate the News Agenda in Election Times. *The International Journal of Press/Politics*, 22(3):275–293, 2017. ISSN 1940-1612. doi: 10.1177/1940161217704969.

Summer Harlow. Framing #ferguson: A comparative analysis of media tweets in the u.s., u.k., spain, and france. *International Communication Gazette*, 81(6-8): 623–643, 2020. doi: 10.1177/1748048518822610. URL https://doi.org/10.1177/1748048518822610.

Summer Harlow and Thomas J. Johnson. The Arab Spring: Overthrowing the Protest Paradigm? How The New York Times, Global Voices and Twitter Covered the

Egyptian Revolution. *International Journal of Communication*, 5(0):16, 2011. ISSN 1932-8036. URL http://ijoc.org/index.php/ijoc/article/download/1239/611.

Summer Harlow, Ramón Salaverría, Danielle K. Kilgo, and Víctor García-Perdomo. Protest Paradigm in Multimedia: Social Media Sharing of Coverage About the Crime of Ayotzinapa, Mexico. *Journal of Communication*, 67(3):328–349, 2017. ISSN 00219916. doi: 10.1111/jcom.12296.

Summer Harlow, Danielle K. Kilgo, Ramón Salaverría, and Víctor García-Perdomo. Is the Whole World Watching? Building a Typology of Protest Coverage on Social Media From Around the World. *Journalism Studies*, 21(11):1590–1608, August 2020. ISSN 1461-670X, 1469-9699. doi: 10.1080/1461670X.2020.1776144.

Christian M. Hennig, Marina Meilă, Fionn Murtagh, and Roberto Rocci, editors. *Handbook of cluster analysis*. Number 9 in Chapman & Hall/CRC handbooks of modern statistical methods. CRC Press, Taylor & Francis Group, Boca Raton, 2016. ISBN 978-1-4665-5188-6.

Edward S. Herman. *Triumph of the market: Essays on economics, politics, and the media*. South End Press, Boston, 1995. ISBN 978-0-89608-521-3.

Edward S. Herman and Noam Chomsky. *Manufacturing consent: The political economy of the mass media*. Pantheon Books, New York, NY, 1988. ISBN 978-0-375-71449-8.

Alastair Hetherington. *News, Newspapers and Television*. Palgrave Macmillan UK, London, 1985. ISBN 978-0-333-38606-4. doi: 10.1007/978-1-349-18000-4.

Paul M. Hirsch. Occupational, organizational and institutional models in mass media research: Toward an integrated framework. In Paul M. Hirsch, Peter V. Miller, and F. Gerald Kline, editors, *Strategies for communication research*, Sage annual reviews of communication research, pages 13–40. Sage, Beverly Hills, 1977. ISBN 978-0-8039-0892-5.

Paul M. Hirsch, Peter V. Miller, and F. Gerald Kline, editors. *Strategies for communication research*, volume 6 of *Sage annual reviews of communication research*. Sage, Beverly Hills, 1977. ISBN 978-0-8039-0892-5.

Peter Hocke. Determining the Selection Bias in Local and National Newspaper Reports on Protest Events. In Dieter Rucht, Ruud Koopmans, and Friedhelm Neidhardt, editors, *Acts of Dissent*. Rowman & Littlefield, Lanham, MD, 1999.

Daniel J. Hopkins and Gary King. A Method of Automated Nonparametric Content Analysis for Social Science. *American Journal of Political Science*, 54(1):229–247, 2010. doi: 10.1111/j.1540-5907.2009.00428.x.

John L. Horn. A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2):179–185, June 1965. ISSN 0033-3123, 1860-0980. doi: 10.1007/BF02289447. URL http://link.springer.com/10.1007/BF02289447.

Simon Hug and Dominique Wisler. Correcting for Selection Bias in Social Movement Research. *Mobilization: An International Journal, 1998, 3(2):*, 1998(3(2)):141–161, 1998. doi: 10.17813/maiq.3.2.6ptv3133154x28n5.

S. Hughes and C. Mellado. Protest and Accountability without the Press: The Press, Politicians, and Civil Society in Chile. *The International Journal of Press/Politics*, 21(1):48–67, 2015. ISSN 1940-1612. doi: 10.1177/1940161215614565.

Edda Humprecht and Florin Büchel. More of the Same or Marketplace of Opinions? A Cross-National Comparison of Diversity in Online News Reporting. *The International Journal of Press/Politics*, 18(4):436–461, 2013. ISSN 1940-1612. doi: 10.1177/1940161213497595.

Swen Hutter and Endre Borbáth. Challenges from left and right: the long-term dynamics of protest and electoral politics in western europe. *European Societies*, 21(4): 487–512, 2019. doi: 10.1080/14616696.2018.1494299. URL https://doi.org/10.1080/14616696.2018.1494299.

Amani Ismail, Gayane Torosyan, and Melissa Tully. Social media, legacy media and gatekeeping: the protest paradigm in news of Ferguson and Charlottesville. *The Communication Review*, 22(3):169–195, July 2019. ISSN 1071-4421, 1547-7487. doi: 10.1080/10714421.2019.1651153. URL https://www.tandfonline.com/doi/full/10.1080/10714421.2019.1651153.

Shanto Iyengar. *Is Anyone Responsible? How Television Frames Political Issues*. American politics and political economy series. University Of Chicago Press, Chicago, 1991. ISBN 0-226-38853-0.

Shanto. Iyengar and Donald R. Kinder. *News that matters: Television and American opinion*. University Of Chicago Press, Chicago, 1987. ISBN 0-226-38856-5.

Carina Jacobi, Wouter van Atteveldt, and Kasper Welbers. Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital Journalism*, 4(1): 89–106, 2015. ISSN 2167-0811. doi: 10.1080/21670811.2015.1093271.

S. Jha. Exploring Internet Influence on the Coverage of Social Protest. Content Analysis Comparing Protest Coverage in 1967 and 1999. *Journalism & Mass Communication Quarterly*, 84(1):40–57, 2007. ISSN 1077-6990. doi: 10.1177/107769900708400104.

I. T. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics. Springer, New York, 2nd ed edition, 2002. ISBN 978-0-387-95442-4.

Andreas Jungherr, Oliver Posegga, and Jisun An. Discursive Power in Contemporary Media Systems: A Comparative Framework. *The International Journal of Press/Politics*, 24(4):404–425, October 2019. ISSN 1940-1612, 1940-1620. doi: 10.1177/1940161219841543. URL http://journals.sagepub.com/doi/10.1177/1940161219841543.

Dan Jurafsky and James H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall Series in Artificial Intelligence. Pearson Prentice Hall, Upper Saddle

River, N.J, 3rd ed draft, december 2020 edition, 2020. ISBN 978-0-13-187321-6. URL https://web.stanford.edu/~jurafsky/slp3/ed3book_dec302020.pdf.

Timothy P. Jurka and Yoshimasa Tsuruoka. *maxent: Low-memory Multinomial Logistic Regression with Support for Text Classification*, 2013. URL https://CRAN.R-project.org/package=maxent. R package version 1.3.3.1.

Marion R. Just. What's News: A View from the Twenty–First Century. In George C. Edwards, III, Lawrence R. Jacobs, and Robert Y. Shapiro, editors, *The Oxford handbook of American public opinion and the media*, The Oxford handbooks of American politics. Oxford University Press, Oxford, UK, 2013. ISBN 0-19-954563-4. doi: 10.1093/oxfordhb/9780199545636.003.0007.

Volha Kananovich. Framing the Taxation-Democratization Link: An Automated Content Analysis of Cross-National Newspaper Data. *The International Journal of Press/Politics*, 23(2):247–267, April 2018. ISSN 1940-1612, 1940-1620. doi: 10.1177/1940161218771893. URL http://journals.sagepub.com/doi/10.1177/1940161218771893.

Danielle K. Kilgo and Summer Harlow. Protests, Media Coverage, and a Hierarchy of Social Struggle. *The International Journal of Press/Politics*, 24(4):508–530, October 2019. ISSN 1940-1612, 1940-1620. doi: 10.1177/1940161219853517. URL http://journals.sagepub.com/doi/10.1177/1940161219853517.

Danielle K. Kilgo, Rachel R. Mourão, and George Sylvie. Martin to Brown: How time and platform impact coverage of the Black Lives Matter movement. *Journalism Practice*, 13(4):413–430, April 2019. ISSN 1751-2786, 1751-2794. doi: 10.1080/17512786.2018.1507680. URL https://www.tandfonline.com/doi/full/10.1080/17512786.2018.1507680.

Kisun Kim and Saif Shahin. Ideological parallelism: toward a transnational understanding of the protest paradigm. *Social Movement Studies*, 19(4):391–407, July

2020. ISSN 1474-2837, 1474-2829. doi: 10.1080/14742837.2019.1681956. URL https://www.tandfonline.com/doi/full/10.1080/14742837.2019.1681956.

Gary King, Robert O. Keohane, and Sidney Verba. *Designing social inquiry: Scientific inference in qualitative research.* Princeton paperbacks. Princeton Univ. Press, Princeton, NJ, [reprinted] edition, 1995. ISBN 0-691-03470-2.

Gary King, Patrick Lam, and Margaret E. Roberts. Computer-Assisted Keyword and Document Set Discovery from Unstructured Text. *American Journal of Political Science*, 61(4):971–988, 2017. doi: 10.1111/ajps.12291. URL http://j.mp/2nxUa8N.

Ronald S. King. *Cluster analysis and data mining: An introduction.* Mercury Learning and Information, Dulles, Virginia; Boston, Massachusetts; New Delhi, January 2015. ISBN 978-1-938549-38-0.

Ruud Koopmans. Movements and media: Selection processes and evolutionary dynamics in the public sphere. *Theory and Society*, 33(3/4):367–391, 2004. ISSN 0304-2421. doi: 10.1023/B:RYSO.0000038603.34963.de.

Hanspeter Kriesi, Edgar Grande, Martin Dolezal, Marc Helbing, Dominic Höglinger, Swen Hutter, and Bruno Wüest. Poldem-protest dataset 6 European countries, 2020a. URL https://poldem.eui.eu/download/protest-events/.

Hanspeter Kriesi, Bruno Wüest, Jasmine Lorenzini, Peter Makarov, Matthias Enggist, Klaus Rothenhäusler, Thomas Kurer, Silja Häusermann, Patrice Wangen, Argyrios Altiparmakis, Endre Borbáth, Björn Bremer, Theresa Gessler, Sophia Hunger, Swen Hutter, Julia Schulte-Cloos, and Chendi Wang. Poldem-protest dataset 30 European countries, 2020b. URL https://poldem.eui.eu/download/protest-events/.

Klaus Krippendorff. *Content analysis: An introduction to its methodology.* 2nd edition edition, 2004. ISBN 0-7619-1545-1.

Thomas B. Ksiazek, Limor Peer, and Kevin Lessard. User engagement with online news: Conceptualizing interactivity and exploring the relationship between online news videos and user comments. *New Media & Society*, 18(3):502–520, 2014. ISSN 1461-4448. doi: 10.1177/1461444814545073.

Raymond Kuhn. *Politics and the media in Britain*. Contemporary political studies. Palgrave Macmillan, Basingstoke and New York, 2007. ISBN 978-0-333-92690-1.

Thomas Samuel Kuhn. *The Structure of Scientific Revolutions*. Chicago Univ. Press, Chicago, Ill. and Chicago, Ill., 2nd ed., enlarged, 3rd impr edition, 1970. ISBN 0-226-45804-0.

Maria Kyriakidou and Jose Javier Olivas Osuna. The Indignados protests in the Spanish and Greek press: Moving beyond the 'protest paradigm'? *European Journal of Communication*, 32(5):457–472, 2017. ISSN 0267-3231. doi: 10.1177/0267323117720342.

Maria Kyriakidou, José Javier Olivas Osuna, and Max Hänska. The Indignados in the European press. In Jeffrey Wimmer, Cornelia Wallner, Rainer Winter, and Karoline Oelsner, editors, *(Mis)understanding political participation: digital practices, new forms of participation and the renewal of democracy*, number 13 in Routledge studies in European communication research and education. Routledge, Taylor & Francis Group, New York London, 2017. ISBN 9781315620596. OCLC: 1019711067.

Pascal D. König. Clusteranalysen. In Claudius Wagemann, Achim Goerres, and Markus Siewert, editors, *Handbuch Methoden der Politikwissenschaft*, Springer Reference Sozialwissenschaften, pages 1–37. Springer Fachmedien, Wiesbaden, 2018. ISBN 978-3-658-16937-4. doi: 10.1007/978-3-658-16937-4_32-1. URL https://doi.org/10.1007/978-3-658-16937-4_32-1.

J. Richard Landis and Gary G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174, 1977. ISSN 0006341X, 15410420. URL http://www.jstor.org/stable/2529310.

Ana Inés Langer and Johannes B. Gruber. Political Agenda Setting in the Hybrid Media System. Why Legacy Media Still Matter a Great Deal. *The International Journal of Press/Politics*, 26(2):313–340, 2021. doi: 10.1177/1940161220925023. URL https://doi.org/10.1177/1940161220925023.

John Langer. *Tabloid television: Popular journalism and the \dqother news\dq*. Communication and society. Routledge, London and New York, 1998. ISBN 978-0-415-06637-2.

Francis L. F. Lee. Triggering the protest paradigm: Examining factors affecting news coverage of protests. *International Journal of Communication*, 8(1):2725–2746, 2014. ISSN 1932-8036.

Thomas J. Leeper. Interpreting Regression Results using Average Marginal Effects with R's margins. R package vignette, 2021a. URL https://cran.r-project.org/web/packages/margins/vignettes/TechnicalDetails.pdf.

Thomas J. Leeper. *margins: Marginal Effects for Model Objects*, 2021b. URL https://github.com/leeper/margins. R package version 0.3.26.

Kalev H. Leetaru. The GDELT Project, 2020. URL https://www.gdeltproject.org/#intro.

Joy Leopold and Myrtle P. Bell. News media and the racialization of protest: an analysis of Black Lives Matter articles. *Equality, Diversity and Inclusion: An International Journal*, 36(8):720–735, January 2017. ISSN 2040-7149. doi: 10.1108/EDI-01-2017-0010. URL https://doi.org/10.1108/EDI-01-2017-0010. Publisher: Emerald Publishing Limited.

Tommy Leung and Nathan Perkins. Count Love project. Protests for a kinder world. URL https://countlove.org/.

Vladimir Levenshtein. Binary codes capable of correcting spurious insertions and deletions of ones. *Soviet Physics-Doklady*, 10(8), 1966. URL https://nymity.ch/sybilhunting/pdf/Levenshtein1966a.pdf.

Michael Lipsky. Protest as a Political Resource. *American Political Science Review*, 62(04):1144–1158, 1968. ISSN 0003-0554. doi: 10.2307/1953909.

Steven Livingston and Lance W. Bennett. Gatekeeping, Indexing, and Live-Event News: Is Technology Altering the Construction of News? *Political Communication*, 20(4):363–380, 2003. doi: 10.1080/10584600390244121. URL https://doi.org/10.1080/10584600390244121.

J. Scott Long. *Regression models for categorical and limited dependent variables*. Number 7 in Advanced quantitative techniques in the social sciences. Sage Publications, Thousand Oaks, 1997. ISBN 978-0-8039-7374-9.

Wiebke Loosen and Jan-Hinrik Schmidt. (Re-)discovering the audience: The relationship between journalism and audience in networked digital media. *Information, Communication & Society*, 15(6):867–887, 2012. ISSN 1369-118X. doi: 10.1080/1369118X.2012.665467.

T. I.M. Loughran and Bill Mcdonald. When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance*, 66(1):35–65, 2011. ISSN 00221082. doi: 10.1111/j.1540-6261.2010.01625.x.

Niklas Luhmann. *The reality of the mass media*. Cultural memory in the present. Stanford University Press, Stanford, Calif, 2000. ISBN 0-8047-4076-3.

Phil MacGregor. Tracking the online audience: Metric data start a subtle revolution. *Journalism Studies*, 8(2):280–298, 2007. ISSN 1461-670X. doi: 10.1080/14616700601148879.

Don Mackay and Nathan Yates. Police hit as hunt yobs hurl fireworks in riot. *Daily Mirror*, 2004, 2004-09-16.

Daniel Maier, A. Waldherr, P. Miltner, G. Wiedemann, A. Niekler, A. Keinert, B. Pfetsch, G. Heyer, U. Reber, T. Häussler, H. Schmid-Petri, and S. Adam. Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology. *Communication Methods and Measures*, 12(2-3):93–118, 2018. ISSN 1931-2458. doi: 10.1080/19312458.2018.1430754.

Christopher D Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to information retrieval*. Cambridge University Press, New York, 2008. ISBN 978-0-511-41405-3.

Geoffrey R. Marczyk, David DeMatteo, and David Festinger. *Essentials of research design and methodology*. Essentials of behavioral science series. John Wiley & Sons, Hoboken, NJ, 2005. ISBN 0-471-47053-8.

Jörg Matthes. What's in a Frame? A Content Analysis of Media Framing Studies in the World's Leading Communication Journals, 1990-2005. *Journalism & Mass Communication Quarterly*, 86(2):349–367, 2009. ISSN 1077-6990. doi: 10.1177/107769900908600206.

Jörg Matthes. Zum Gehalt der Framing-Forschung: Eine kritische Bestandsaufnahme. In Frank Marcinkowski, editor, *Framing als politischer Prozess*, Schriftenreihe Politische Kommunikation und demokratische Öffentlichkeit, pages 17–28. Nomos, Baden-Baden, 2013. ISBN 978-3-8487-0238-1.

Jörg Matthes and Matthias Kohring. The Content Analysis of Media Frames: Toward Improving Reliability and Validity. *Journal of Communication*, 58(2):258–279, 2008. ISSN 00219916. doi: 10.1111/j.1460-2466.2008.00384.x.

Jörg Matthes and Christian Schemer. Diachronic Framing Effects in Competitive Opinion Environments. *Political Communication*, 29(3):319–339, 2012. ISSN 1058-4609. doi: 10.1080/10584609.2012.694985.

Doug McAdam, John McCarthy, Susan Olzak, and Sarah Soule. Dynamics of collective action. URL https://web.stanford.edu/group/collectiveaction/cgi-bin/drupal/.

John D. McCarthy, Clark McPhail, and Jackie Smith. Images of Protest: Dimensions of Selection Bias in Media Coverage of Washington Demonstrations, 1982 and 1991. *American Sociological Review*, 61(3):478–499, 1996. ISSN 0003-1224. doi: 10.2307/2096360. URL http://www.jstor.org/stable/2096360.

John D. McCarthy, Clark McPhail, Jackie Smith, and Louis Crishock. Electronic and Print Media Representations of Washington D.C. Demonstrations, 1982 and 1991. In Dieter Rucht, Ruud Koopmans, and Friedhelm Neidhardt, editors, *Acts of Dissent*, pages 113–130. Rowman & Littlefield, Lanham, MD, 1999. URL http://d-scholarship.pitt.edu/20703/.

Maxwell E. McCombs and Donald L. Shaw. The Agenda-Setting Function of Mass Media. *Public Opinion Quarterly*, 36(2):176, 1972. ISSN 0033362X. doi: 10.1086/267990.

Patrick McCurdy. Social Movements, Protest and Mainstream Media. *Sociology Compass*, 6(3):244–255, 2012. ISSN 17519020. doi: 10.1111/j.1751-9020.2011.00448.x.

Patrick M. McCurdy. Inside the media event: Examining the media practices of Dissent! at the Hori-Zone eco-village at the 2005 G8 Gleneagles Summit. *Communications*, 33(3), January 2008. ISSN 0341-2059, 1613-4087. doi: 10.1515/COMM.2008.019. URL https://www.degruyter.com/doi/10.1515/COMM.2008.019.

Thomas McFarlane and Iain Hay. The battle for Seattle: Protest and popular geopolitics in The Australian newspaper. *Political Geography*, 22(2):211–232, 2003. ISSN 09626298. doi: 10.1016/S0962-6298(02)00090-2.

George McKay, editor. *DiY Culture: Party & protest in nineties britain*. Verso, London and New York, 1998. ISBN 978-1-85984-260-7.

David McKnight. A change in the climate? The journalism of opinion at News Corporation. *Journalism*, 11(6):693–706, December 2010. ISSN 1464-8849, 1741-3001. doi: 10.1177/1464884910379704. URL http://journals.sagepub.com/doi/10.1177/1464884910379704.

Doug McLeod. Social Protest, 2011. URL http://www.oxfordbibliographies.com/view/document/obo-9780199756841/obo-9780199756841-0005.xml. Accessed: 2017-09-20.

Douglas M. McLeod. Communicating deviance: The effects of television news coverage of social protest. *Journal of Broadcasting & Electronic Media*, 39(1):4–19, 1995. ISSN 0883-8151. doi: 10.1080/08838159509364285.

Douglas M. McLeod. The protest paradigm and news coverage of the 'right to party' movement. In David A. Schultz, editor, *\dqIt's show time!*, pages 29–49. 2000.

Douglas M. McLeod. News Coverage and Social Protest: How the Media's Protect Paradigm Exacerbates Social Conflict. *J. Disp. Resol.*, pages 185–194, 2007. URL http://scholarship.law.missouri.edu/jdr/vol2007/iss1/12/.

Douglas M. McLeod and B. H. Detenber. Framing effects of television news coverage of social protest. *Journal of Communication*, 49(3):3–23, 1999. ISSN 0021-9916. doi: 10.1111/j.1460-2466.1999.tb02802.x.

Douglas M. McLeod and J. K. Hertog. The Manufacture of 'Public Opinion' by Reporters: Informal Cues for Public Perceptions of Protest Groups. *Discourse & Society*, 3(3):259–275, 1992. ISSN 0957-9265. doi: 10.1177/0957926592003003001.

Douglas M. McLeod and J. K. Hertog. Social control, social change and the mass media's role in the regulation of protest groups. In David Demers and K. Viswanath, editors, *Mass media, social control, and social change*, pages 305–330. Iowa State Univ. Pr, Ames, 1999. ISBN 0-8138-2682-9.

Brian McNair. *Journalism and democracy: An evaluation of the political public sphere*. Routledge, London and New York, 2006. ISBN 0-415-21279-0.

Media Reform Coalition. Report: Who Owns the UK Media?, March 2021. URL https://www.mediareform.org.uk/wp-content/uploads/2021/03/Who-Owns-the-UK-Media_final2.pdf.

Robbie Millen. March for anything, anywhere. *The Times*, 2003, 2003-02-12.

M. Mark Miller. Frame Mapping and Analysis of News Coverage of Contentious Issues. *Social Science Computer Review*, 15(4):367–378, 1997. ISSN 0894-4393. doi: 10.1177/089443939701500403.

Glenn W. Milligan and Martha C. Cooper. An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, 50(2):159–179, January 1985. ISSN 0033-3123. doi: 10.1007/BF02294245.

Kirsty Milne. *Manufacturing dissent: Single-issue protest, the public and the press*. Demos, London, 2005. ISBN 1-84180-141-0.

Boris Mirkin. Quadratic error and k-means. In Christian M. Hennig, Marina Meilă, Fionn Murtagh, and Roberto Rocci, editors, *Handbook of cluster analysis*, Chapman & Hall/CRC handbooks of modern statistical methods, pages 33–52. CRC Press, Taylor & Francis Group, Boca Raton, 2016. ISBN 978-1-4665-5188-6. doi: 10.1201/b19706.

Mario Molina and Filiz Garip. Machine Learning for Sociology. *Annual Review of Sociology*, 45(1):27–45, July 2019. ISSN 0360-0572, 1545-2115. doi: 10.1146/annurev-soc-073117-041106. URL https://www.annualreviews.org/doi/10.1146/annurev-soc-073117-041106.

Rachel R. Mourão. From Mass to Elite Protests: News Coverage and the Evolution of Antigovernment Demonstrations in Brazil. *Mass Communication and Society*, 22(1):49–71, January 2019. ISSN 1520-5436, 1532-7825. doi: 10.1080/15205436.2018.1498899. URL https://www.tandfonline.com/doi/full/10.1080/15205436.2018.1498899.

Rachel R. Mourão and Weiyue Chen. Covering Protests on Twitter: The Influences on Journalists' Social Media Portrayals of Left- and Right-Leaning Demonstrations in Brazil. *The International Journal of Press/Politics*, 2(4):194016121988265, January 2019. ISSN 1940-1612. doi: 10.1177/1940161219882653.

Carol Mueller. International Press Coverage of East German Protest Events, 1989. *American Sociological Review*, 62(5):820–832, 1997. ISSN 0003-1224. doi: 10.2307/2657362.

Graham Murdock. Political deviance: the press presentation of a militant mass demonstration. In Stanley Cohen and Jock Young, editors, *The manufacture of news*, Communication and society, pages 156–175. Constable, London, 1973. ISBN 978-0-09-459490-6.

Graham Murdock. Large corporations and the control of the communications industries. In Michael Gurevitch, Tony Bennett, James Curran, and Janet Woollacott, editors, *Culture, society, and the media*, pages 114–147. Routledge, London and New York, 1988. ISBN 0-203-97809-9.

Ralph Negrine. Demonstrations, Protest, and Communication: Changing Media Landscapes—Changing Media Practices? In Kathrin Fahlenbrach, Erling Sivertsen, and Rolf Werenskjold, editors, *Media and revolt*, Protest, culture and society, pages 59–74. Berghahn, New York and Oxford, 2016. ISBN 978-0-85745-999-2.

Laura K. Nelson, Derek Burk, Marcel Knudsen, and Leslie McCall. The Future of Coding: A Comparison of Hand-Coding and Three Types of Computer-Assisted Text Analysis Methods. *Sociological Methods & Research*, 50(1):202–237, February 2021. ISSN 0049-1241, 1552-8294. doi: 10.1177/0049124118769114. URL http://journals.sagepub.com/doi/10.1177/0049124118769114.

Tom Nicholls and Pepper D. Culpepper. Computational Identification of Media Frames: Strengths, Weaknesses, and Opportunities. *Political Communication*, pages 1–23, September 2020. ISSN 1058-4609, 1091-7675. doi: 10.1080/10584609.2020.1812777. URL https://www.tandfonline.com/doi/full/10.1080/10584609.2020.1812777.

Pippa Norris. *Democratic Phoenix: Reinventing Political Activism.* Cambridge University Press, 2002. ISBN 978-0521010535.

Pippa Norris, Stefaan Walgrave, and Peter van Aelst. Who Demonstrates? Antistate Rebels, Conventional Participants, or Everyone? *Comparative Politics*, 37(2):189, 2005. ISSN 00104159. doi: 10.2307/20072882.

Rory O'Connor. Facebook and Twitter Are Reshaping Journalism As We Know It: The rise of Facebook and Twitter herald changes for journalism, and pose serious challenges to about journalistic credibility and trust, January 2009. URL https://www.alternet.org/story/121211/facebook_and_twitter_are_reshaping_journalism_as_we_know_it.

Pamela E. Oliver and Gregory M. Maney. Political Processes and Local Newspaper Coverage of Protest Events: From Selection Bias to Triadic Interactions. *American Journal of Sociology*, 106(2):463–505, 2000. ISSN 0002-9602. doi: 10.1086/316964.

Pamela E. Oliver and Daniel J. Meyer. How Events Enter the Public Sphere: Conflict, Location, and Sponsorship in Local Newspaper Coverage of Public Events. *American Journal of Sociology*, 105(1):38–87, 1999. ISSN 0002-9602. doi: 10.1086/210267.

Karl-Dieter Opp. *Theories of political protest and social movements: A multidisciplinary introduction, critique, and synthesis.* Routledge, London, transferred to digital printing edition, 2015. ISBN 0-415-48388-3.

David Ortiz, Daniel Myers, Eugene Walls, and Maria-Elena Diaz. Where Do We Stand with Newspaper Data? *Mobilization: An International Quarterly*, 10(3):397–419, 2005. ISSN 1086-671X. doi: 10.17813/maiq.10.3.8360r760k3277t42.

M. Oz. Mainstream medias coverage of the Gezi protests and protesters perception of mainstream media. *Global Media and Communication*, 12(2):177–192, 2016. ISSN 1742-7665. doi: 10.1177/1742766516653164.

Zhongdang Pan and Gerald Kosicki. Framing analysis: An approach to news discourse. *Political Communication*, 10(1):55–75, 1993. ISSN 1058-4609. doi: 10.1080/10584609.1993.9962963.

Tao Papaioannou. Participation and Media| Dominant and Emerging News Frames in Protest Coverage: The 2013 Cypriot Anti-Austerity Protests in National Media. *International Journal of Communication*, 14(0), 2020. ISSN 1932-8036. URL https://ijoc.org/index.php/ijoc/article/view/9002.

John H. Parmelee. The agenda-building function of political tweets. *New Media & Society*, 16(3):434–450, 2013. ISSN 1461-4448. doi: 10.1177/1461444813487955.

Thomas B. Pepinsky. The return of the single-country study. *Annual Review of Political Science*, 22(1):187–203, 2019. doi: 10.1146/annurev-polisci-051017-113314. URL https://doi.org/10.1146/annurev-polisci-051017-113314.

Andrea Peters and Torsten Hothorn. *ipred: Improved Predictors*, 2019. URL https://CRAN.R-project.org/package=ipred. R package version 0.9-9.

Jeremy W. Peters. Some newspapers shift coverage after tracking readers online, January 2010. URL http://www.nytimes.com/2010/09/06/business/media/06track.html.

Mark Steven Poster. *What's the matter with the internet?* University of Minnesota Press, Minneapolis, 2001. ISBN 978-0-8166-3835-2.

Martin J. Power, Amanda Haynes, and Eoin Devereux. Reasonable People vs. The Sinister Fringe: Interrogating the framing of Ireland's water charge protestors through the media politics of dissent. *Critical Discourse Studies*, 13(3):261–277, May 2016. ISSN 1740-5904, 1740-5912. doi: 10.1080/17405904.2016.1141694. URL http://www.tandfonline.com/doi/full/10.1080/17405904.2016.1141694.

Jeremy Pressman and Erica Chenoweth. Crowd Counting Consortium. A public interest and scholarly project to document crowds and contention in the united states. URL https://sites.google.com/view/crowdcountingconsortium.

V. Price, D. Tewksbury, and E. Powers. Switching Trains of Thought: The Impact of News Frames on Readers' Cognitive Responses. *Communication Research*, 24(5): 481–506, 1997. ISSN 0093-6502. doi: 10.1177/009365097024005002.

Quinn, Gary, and Brian Trench. Online News Media and Their Audiences, January 2002. URL https://web.archive.org/web/20030826072104/http://www.mudia.org/results/WP1%20Del%201.3%20Web%20Version.pdf.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021. URL https://www.R-project.org/.

Carsten Reinemann, James Stanyer, Sebastian Scherr, and Guido Legnante. Hard and soft news: A review of concepts, operationalizations and key findings. *Journalism: Theory, Practice & Criticism*, 13(2):221–239, February 2012. ISSN 1464-8849, 1741-3001. doi: 10.1177/1464884911427803. URL http://journals.sagepub.com/doi/10.1177/1464884911427803.

Carsten Reinemann, James Stanyer, and Sebastian Scherr. Hard and soft news. In C. H. de Vreese, Frank Esser, and David Nicolas Hopmann, editors, *Comparing political journalism*, pages 131–149. Routledge, London New York, 2016. ISBN 978-1-138-65585-0 978-1-138-65586-7.

R. Reul, S. Paulussen, D. Raeijmaekers, L. van der Steen, and P. Maeseele. Professional journalistic routines and the protest paradigm: The Big Potato Swap in traditional and alternative media. *Journalism*, 2016. ISSN 1464-8849. doi: 10.1177/1464884916636170.

William Revelle. *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois, 2019. URL https://CRAN.R-project.org/package=psych. R package version 1.9.12.

Brian Ripley. *tree: Classification and Regression Trees*, 2019. URL https://CRAN.R-project.org/package=tree. R package version 1.0-40.

Ingrid Rogstad. Is Twitter just rehashing? Intermedia agenda setting between Twitter and mainstream media. *Journal of Information Technology & Politics*, 13(2):142–158, 2016. ISSN 1933-1681. doi: 10.1080/19331681.2016.1160263.

Deana A. Rohlinger. Persuasion and Non-Party Groups in the Digital Age. In Elizabeth Suhay, Bernard Grofman, and Alexander H. Trechsel, editors, *The Oxford Handbook of Electoral Persuasion*, pages 320–339. Oxford University Press, June 2020. ISBN 978-0-19-086080-6. doi: 10.1093/oxfordhb/9780190860806.013.40. URL https://oxfordhandbooks.com/view/10.1093/oxfordhb/9780190860806.001.0001/oxfordhb-9780190860806-e-40.

Michael Rosie and Hugo Gorringe. 'The Anarchists' World Cup': Respectable Protest and Media Panics. *Social Movement Studies*, 8(1):35–53, 2009. ISSN 1474-2837. doi: 10.1080/14742830802591135.

Yves Rosseel. lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2):1–36, 2012. URL http://www.jstatsoft.org/v48/i02/.

Dieter Rucht. Protest Movements and their Media Usages. In Bart Cammaerts, Alice Mattoni, and Patrick McCurdy, editors, *Mediation and Protest Movements*, pages 249–268. Intellect, 2013.

W. Russell Neuman, Lauren Guggenheim, S. Mo Jang, and Soo Young Bae. The Dynamics of Public Attention: Agenda-Setting Theory Meets Big Data. *Journal of Communication*, 64(2):193–214, 2014. ISSN 00219916. doi: 10.1111/jcom.12088.

Charlotte Ryan. *Prime time activism: Media strategies for grassroots organizing*. South End Press, Boston, MA, 1st ed. edition, 1991. ISBN 978-0-89608-402-5.

Idean Salehyan, Cullen S. Hendrix, Jesse Hamner, Christina Case, Christopher Linebarger, Emily Stull, and Jennifer Williams. Social Conflict in Africa: A New Database. *International Interactions*, 38(4):503–511, September 2012. ISSN 0305-0629, 1547-7444. doi: 10.1080/03050629.2012.697426.

David Sanders, Harold Clarke, Marianne Stewart, and Paul Whiteley. The Dynamics of Protest in Britain, 2000-2002. *Parliamentary Affairs*, 56(4):687–699, 2003. ISSN 0031-2290. doi: 10.1093/pa/gsg110.

Michael Scharkow. Thematic content analysis using supervised machine learning: An empirical evaluation using German online news. *Quality & Quantity*, 47(2):761–773, 2013. ISSN 0033-5177. doi: 10.1007/s11135-011-9545-7.

C. Schemer, W. Wirth, and Jörg Matthes. Value Resonance and Value Framing Effects on Voting Intentions in Direct-Democratic Campaigns. *American Behavioral Scientist*, 56(3):334–352, 2012. ISSN 0002-7642. doi: 10.1177/0002764211426329.

D. A. Scheufele. Framing as a theory of media effects. *Journal of Communication*, 49 (1):103–122, 1999. ISSN 0021-9916. doi: 10.1111/j.1460-2466.1999.tb02784.x.

Philip Schlesinger. *Putting \dqreality\dq together: BBC News*. Methuen, London [etc.], 1987. ISBN 978-0-416-90190-0.

Michael Schudson. The Sociology of News Production Revisited. In James Curran and Michael Gurevitch, editors, *Mass media and society*, pages 141–159. Arnold, London, 1997. ISBN 0-340-61418-8.

Revital Sela-Shayovitz and Badi Hasisi. Public order events in the headlines: The media construction of threat and dangerousness of public order events in Israel. *International Journal of Police Science & Management*, 13(4):297–311, 2011. ISSN 1461-3557. doi: 10.1350/ijps.2011.13.4.249.

HA Semetko and PM Valkenburg. Framing European politics: a content analysis of press and television news. *Journal of Communication*, 50(2):93–109, 2000. doi: 10.1111/j.1460-2466.2000.tb02843.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1460-2466.2000.tb02843.x.

Julie Sevenans and Rens Vliegenthart. Political Agenda-Setting in Belgium and the Netherlands: The Moderating Role of Conflict Framing. *Journalism & Mass Communication Quarterly*, 93(1):187–203, March 2016. ISSN 1077-6990, 2161-430X. doi: 10.1177/1077699015607336. URL http://journals.sagepub.com/doi/10.1177/1077699015607336.

S. Shahin, P. Zheng, H. A. Sturm, and D. Fadnis. Protesting the Paradigm: A Comparative Study of News Coverage of Protests in Brazil, China, and India. *The International Journal of Press/Politics*, 21(2):143–164, 2016. ISSN 1940-1612. doi: 10.1177/1940161216631114.

Clay Shirky. *Here comes everybody: The power of organizing without organizations*. Penguin Books, London [u.a.], 2009. ISBN 978-0-7139-9989-1.

P. J. Shoemaker. Media Treatment of Deviant Political Groups. *Journalism & Mass Communication Quarterly*, 61(1):66–82, 1984. ISSN 1077-6990. doi: 10.1177/107769908406100109.

Pamela J. Shoemaker and Stephen D. Reese. *Mediating the message in the 21st century: A media Sociology perspective*. Routledge/Taylor & Francis Group, New York, third edition edition, 2014. ISBN 978-0-415-98914-5.

Doron Shultziner and Aya Shoshan. A Journalists' Protest? Personal Identification and Journalistic Activism in the Israel Social Justice Protest Movement. *The International Journal of Press/Politics*, 23(1):44–69, January 2018. ISSN 1940-1612, 1940-1620. doi: 10.1177/1940161217736889. URL http://journals.sagepub.com/doi/10.1177/1940161217736889.

Jane B. Singer. User-generated visibility: Secondary gatekeeping in a shared media space. *New Media & Society*, 16(1):55–73, 2014. ISSN 1461-4448. doi: 10.1177/1461444813477833.

Jackie Smith, John D. McCarthy, Clark McPhail, and Boguslaw Augustyn. From Protest to Agenda Building: Description Bias in Media Coverage of Protest Events in Washington, D.C. *Social Forces*, 79(4):1397–1423, 2001. ISSN 00377732, 15347605.

Paul M. Sniderman and Sean M. Theriault. The Structure of Political Argument and the Logic of Issue Framing. In Willem E. Saris and Paul M. Sniderman, editors, *Studies in public opinion*, pages 133–165. Princeton University Press, Princeton, N.J. and Oxford, 2004. ISBN 978-0-691-11903-8.

Brian Solis. Is Twitter The CNN Of The New Media Generation?, June 2009. URL https://techcrunch.com/2009/06/17/is-twitter-the-cnn-of-the-new-media-generation/.

Keith Soothill and Chris Grover. A Note on Computer Searches of Newspapers. *Sociology*, 31(3):591–596, 1997. ISSN 00380385, 14698684.

Bartholomew H. Sparrow. *Uncertain guardians: the news media as a political institution*, volume 1999: 1 of *Interpreting American politics*. Johns Hopkins University Press, Baltimore, Md, 1999. ISBN 978-0-8018-6036-2.

Lia-Paschalia Spyridou. Producing protest news: Representations of contentious collective actions in mainstream print media. *The Cyprus Review*, (27(1)):71–105, 2015.

James Stanyer. Web 2.0 and the transformation of news and journalism. In Andrew Chadwick and Philip N. Howard, editors, *Routledge handbook of internet politics*, Routledge Handbooks, pages 201–. Routledge, London and New York, 2010. ISBN 978-0-415-42914-6.

John Street. *Mass media, politics and democracy.* Palgrave Macmillan, Houndmills, Basingstoke, Hampshire and New York, 2nd ed. edition, 2011. ISBN 978-1-4039-4734-5.

The Sun. St paul's stays shut as protest continues. *The Sun*, 2011, 2011-10-24.

K. Tenenboim-Weinblatt. Producing Protest News: An Inquiry into Journalists' Narratives. *The International Journal of Press/Politics*, 19(4):410–429, 2014. ISSN 1940-1612. doi: 10.1177/1940161214540941.

Yannis Theocharis, Will Lowe, Jan W. van Deth, and Gema García-Albacete. Using Twitter to mobilize protest action: online mobilization patterns and action repertoires in the Occupy Wall Street, Indignados, and Aganaktismenoi movements. *Information, Communication & Society*, 18(2):202–220, 2014. ISSN 1369-118X. doi: 10.1080/1369118X.2014.948035.

Thesaurus.com. *Synonyms of protest.* 2018. URL http://www.thesaurus.com/browse/protest.

Charles Tilly. *Popular contention in Britain: 1758-1834.* Harvard University Press, Cambridge Mass, 1995. ISBN 978-0-674-68980-0.

Charles Tilly. Event catalogs as theories. *Sociological Theory*, 20(2):248–254, 2002. ISSN 07352751. URL http://www.jstor.org/stable/3108648.

Charles Tilly. *Contentious performances.* Cambridge studies in contentious politics. Cambridge University Press, Cambridge and New York, second edition edition, 2012. ISBN 978-0-521-51584-9.

Ilija Tomanić Trivundža and Sašo Slaček Brlek. Looking for Mr Hyde: The protest paradigm, violence and (de)legitimation of mass political protests. *International Journal of Media & Cultural Politics*, 13(1):131–148, March 2017. ISSN 1740-8296. doi: 10.1386/macp.13.1-2.131_1. URL http://www.ingentaconnect.com/content/10.1386/macp.13.1-2.131_1.

Gaye Tuchman. Objectivity as Strategic Ritual: An Examination of Newsmen's Notions of Objectivity. *American Journal of Sociology*, 77(4):660–679, 1972. ISSN 0002-9602.

Gaye Tuchman. Making News by Doing Work: Routinizing the Unexpected. *American Journal of Sociology*, 79(1):110–131, 1973. ISSN 0002-9602.

Ralph H. Turner. The Public Perception of Protest. *American Sociological Review*, 34 (6):815–831, 1969. ISSN 0003-1224. doi: 10.2307/2095975.

Jarek Tuszynski. *caTools: Tools: Moving Window Statistics, GIF, Base64, ROC AUC, etc*, 2020. URL https://CRAN.R-project.org/package=caTools. R package version 1.18.0.

Janani Umamaheswar. Policing and Racial (In)Justice in the Media: Newspaper Portrayals of the "Black Lives Matter" Movement. *Civic Sociology*, 1(1):12143, March 2020. ISSN 2637-9155. doi: 10.1525/001c.12143. URL https://online.ucpress.edu/cs/article/doi/10.1525/001c.12143/112918/Policing-and-Racial-InJustice-in-the-Media.

Peter van Aelst and Stefaan Walgrave. Who is that (wo)man in the street? From the normalisation of protest to the normalisation of the protester. *European Journal of Political Research*, 39(4):461–486, 2001. ISSN 0304-4130. doi: 10.1111/1475-6765.00582.

Peter van Aelst, Jesper Strömbäck, Toril Aalberg, Frank Esser, Claes Vreese, Jörg Matthes, David Hopmann, Susana Salgado, Nicolas Hubé, Agnieszka Stępińska, Stylianos Papathanassopoulos, Rosa Berganza, Guido Legnante, Carsten Reinemann, Tamir Sheafer, and James Stanyer. Political communication in a high-choice media environment: a challenge for democracy? *Annals of the International Communication Association*, 41(1):3–27, January 2017. ISSN 2380-8985. doi: 10.1080/23808985.2017.1288551.

M.P.J. van der Loo. The stringdist package for approximate string matching. *The R Journal*, 6:111–122, 2014. URL https://CRAN.R-project.org/package=stringdist.

B. van Gorp. Where is the Frame?: Victims and Intruders in the Belgian Press Coverage of the Asylum Issue. *European Journal of Communication*, 20(4):484–507, 2005. ISSN 0267-3231. doi: 10.1177/0267323105058253.

Baldwin van Gorp. The Constructionist Approach to Framing: Bringing Culture Back In. *Journal of Communication*, 57(1):60–78, 2007. ISSN 00219916. doi: 10.1111/j.0021-9916.2007.00329.x.

Anastasia Veneti, Achilleas Karadimitriou, and Stamatis Poulakidakos. Media Ecology and the Politics of Dissent: Representations of the Hong Kong Protests in *The Guardian* and *China Daily*. *Social Media + Society*, 2(3):205630511666217, September 2016. ISSN 2056-3051, 2056-3051. doi: 10.1177/2056305116662175. URL http://journals.sagepub.com/doi/10.1177/2056305116662175.

Hong Tien Vu. The online audience as gatekeeper: The influence of reader metrics on news editorial selection. *Journalism: Theory, Practice & Criticism*, 15(8):1094–1110, 2013. ISSN 1464-8849. doi: 10.1177/1464884913504259.

Karin Wahl-Jorgensen. The construction of the public in letters to the editor. *Journalism: Theory, Practice & Criticism*, 3(2):183–204, 2002. ISSN 1464-8849. doi: 10.1177/146488490200300203.

Omar Wasow. Agenda seeding: How 1960s black protests moved elites, public opinion and voting. *American Political Science Review*, 114(3):638–659, 2020. ISSN 0003-0554. doi: 10.1017/S000305542000009X.

Kohei Watanabe. *Newsmap*, 2018. URL https://doi.org/10.1080/21670811.2017.1293487.

David A. Weaver and Bruce Bimber. Finding News Stories: A Comparison of Searches Using Lexisnexis and Google News. *Journalism & Mass Communication Quarterly*, 85(3):515–530, 2008. ISSN 1077-6990. doi: 10.1177/107769900808500303.

David A. Weaver and Joshua M. Scacco. Revisiting the Protest Paradigm: The Tea Party as Filtered through Prime-Time Cable News. *The International Journal of Press/Politics*, 18(1):61–84, 2012. ISSN 1940-1612. doi: 10.1177/1940161212462872.

N. B. Weidmann and E. G. Rød. The internet and political protest in autocracies, 2019. URL https://mmadatabase.org/about/documentation/.

Gabriel Weimann and Hans-Bernd Brosius. A New Agenda for Agenda-Setting Research in the Digital Era. In Gerhard Vowe and Philipp Henn, editors, *Political communication in the online world*, Routledge research in political communication, pages 26–44. Routledge, New York and London, 2016. ISBN 978-1-315-70749-5.

Kasper Welbers, Wouter van Atteveldt, and Kenneth Benoit. Text Analysis in R. *Communication Methods and Measures*, 11(4):245–265, 2017. ISSN 1931-2458. doi: 10.1080/19312458.2017.1387238.

Wikipedia. Protest, January 2020. URL https://en.wikipedia.org/wiki/Protest#Forms.

Gadi Wolfsfeld. *Media and political conflict: News from the Middle East*. Cambridge University Press, Cambridge [England] and New York, NY, USA, 1997. ISBN 978-0-521-58967-3.

Gadi Wolfsfeld. *Making sense of media and politics: Five principles in political communication*. Routledge, New York, 2011. ISBN 978-0-203-83987-4. URL http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10514336.

Gadi Wolfsfeld, Elad Segev, and Tamir Sheafer. Social Media and the Arab Spring. *The International Journal of Press/Politics*, 18(2):115–137, 2013. ISSN 1940-1612. doi: 10.1177/1940161212471716.

Ruud Wouters. Patterns in Advocacy Group Portrayal: Comparing Attributes of Protest and Non-Protest News Items Across Advocacy Groups. *Journalism & Mass Communication Quarterly*, 92(4):898–914, 2015a. ISSN 1077-6990. doi: 10.1177/1077699015596327.

Ruud Wouters. Reporting Demonstrations: On Episodic and Thematic Coverage of Protest Events in Belgian Television News. *Political Communication*, 32(3):475–496, 2015b. ISSN 1058-4609. doi: 10.1080/10584609.2014.958257.

Marvin N. Wright and Andreas Ziegler. ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software*, 77 (1):1–17, 2017. doi: 10.18637/jss.v077.i01.

Kaibin Xu. Framing Occupy Wall Street: A Content Analysis of The New York Times and USA Today. *International Journal of Communication; Vol 7 (2013)*, 2013.

Dannagal G. Young. Political satire and occupy wall street: How comics co-opted strategies of the protest paradigm to legitimize a movement. *International Journal of Communication*, 7(1):371–393, 2013. ISSN 1932-8036.

Lori Young and Stuart Soroka. Affective News: The Automated Coding of Sentiment in Political Texts. *Political Communication*, 29(2):205–231, April 2012. ISSN 1058-4609, 1091-7675. doi: 10.1080/10584609.2012.671234. URL http://www.tandfonline.com/doi/abs/10.1080/10584609.2012.671234.

J. Zaller. *A Theory of Media Politics: How the Interests of Politicians, Journalists, and Citizens Shape the News.* University Of Chicago Press, 1999.

Conrad Ziller. Multiple Regression mit voneinander abhängigen Beobachtungen: Random-Effects und Fixed-Effects. In Claudius Wagemann, Achim Goerres, and Markus Siewert, editors, *Handbuch Methoden der Politikwissenschaft*, Springer Reference Sozialwissenschaften, pages 1–27. Springer Fachmedien, Wiesbaden, 2018. ISBN 978-3-658-16937-4. doi: 10.1007/978-3-658-16937-4_28-2. URL https://doi.org/10.1007/978-3-658-16937-4_32-1.

Appendices

A Extended Codebook and Coding Instructions

Purpose and Background

The purpose of this extended codebook is to make the employed coding scheme as clear as possible for coders on the project, document the coding decisions I have made and to present the approach to others in the hope they can make use of it.

Since frames are an implicit construct, coding them holistically can decrease reliability and validity as both frame detection and document classification involves a high degree of subjective judgement by coders. This is especially true if coders have to deal with longer texts that describe a topic not just from one but, in an attempt to appear objective, from multiple angles. Reliability can suffer in this case as some coders focus more on one part of the text while others find another part more important and therefore assign a different frame.

As background for the coding it should be known that frames are understood here to be different combinations of certain frame elements (*Problem Definition, Causal Attribution, Moral Evaluation* and *Treatment recommendation*) based on the canonical definition by Entman (1993).

To find these combinations, and therefore the employed frames, frame elements are coded separately and in each paragraph before determining the frame of a newspaper article empirically. Specifically, the goal is to determine for each paragraph what the *Problem Definition*, *Causal Attribution, Moral Evaluation* and *Treatment recommendation* are employed. This procedure relies on smaller, more transparent human decisions as frame elements are more explicit and easier to code than holistic frames (i.e. trying to determine the frame for the entire text at once). As these frame elements are somewhat abstract, they are further divided into variables for content analysis. *Problem Definition* is determined by coding the variable Topic and Actor, *Causal Attribution* is determined by the variables Benefit and one Risk Attribution*, Moral Evaluation* by Benefit and one Risk and *Treatment recommendation* by positive or negative Judgement – further details about this follow in the respective sections below.

Coding Instructions

During initial coding, variables were identified and coded at the same time. To make things easier for additional coders, coding was split into two separate tasks for this stage: identification of codes and classification. The coding sample contains two different files. One called “4.3.4._clss_sample” (the classification sample) and one called “4.3.4._idd_sample.xlsx” (the identification sample).

Please **do the classification first** as this will help you to familiarize yourself with the existing codes. For row in the document, check the variable column first to see which variable you are dealing with. See below for the different variable which codes are available. Read the paragraph in the text column and assign the code you think fits the paragraph best. The paragraphs were sampled from the paragraphs which contain a code so you should be able to find a code in each text. More on what to choose if the decision is ambiguous can be found below in the description of the codes. An overview of the codes can be found in the table “Variables and Codes for manual content analysis“ and a more detailed description of the codes can be found below. You can ignore the additional columns or consult them if you think the headline, for example, might be helpful in determining the code-

After the classification task is done, look at the second file with the **identification sample**. This task should be a lot quicker. Simple write 1 or TRUE into the code column if you think this variable exists in the paragraph or 0 or FALSE if you think there is no Topic, Actor etc.

Overview

Purpose and Background 2

Coding Instructions 3

Overview 4

1. Problem Definition 7

  1.1. Topic 7

    1.1.1. Clash 7

    1.1.2. Violence/Crime 8

    1.1.3. Cause 8

    1.1.4. Protesters 9

    1.1.5. Spectacle 9

    1.1.6. Protest tactics 10

    1.1.7. Policing tactics 10

    1.1.8. Response 11

    1.1.9. Confrontation/Showdown 11

    1.1.10. Public opinion 12

    1.1.11. Nuisance 12

    1.1.12. Other activism 12

    1.1.13. judicial prosecution 13

    1.1.14. Effect of protest 13

    1.1.15. Event 14

  1.2. Actor 14

    1.2.1. Protesters 14

    1.2.2. Police 14

    1.2.3. Officials 14

    1.2.4. Business 14

    1.2.5. Other Political Elite 14

1.2.6. Other 15
2. Causal Attribution 16
2.1. Benefit Attribution 16
2.1.1. Police 16
2.1.2. Protesters 17
2.1.3. Officials 17
2.2. Risk Attribution 17
2.2.1. Protesters 17
2.2.2. Police 18
2.2.3. Business 18
2.2.4. Officials 18
2.2.5. Media 18
2.2.6. Other Pol. Elite 19
2.2.7. Other 19
3. Moral Evaluation 20
3.1. Benefit 20
3.1.1. Reinstating public order 20
3.1.2. (Just) Cause 20
3.1.3. Initiated public debate 20
3.2. Risk 21
3.2.1. Grievance 21
3.2.2. Public safety 21
3.2.3. Property destruction 22
3.2.4. Harassment 22
3.2.5. Breaking laws 23
3.2.6. Decay of morals or other social norms 23
3.2.7. Nuisance 23

3.2.8. Harm discussion/cause ... 24
3.2.9. Trivializing (political) discussion ... 24
3.2.10. Bad for business ... 25
3.2.11. Costs of demonstrations ... 25
3.2.12. Suppression/Censorship ... 26
4. Treatment ... 27
4.1. Judgement ... 27
4.1.1. Negative ... 27
4.1.2. Positive ... 28

1. Problem Definition

For ***Problem Definition*** the idea was that this frame element can be conceived by coding the *topic*, or central issue that is discussed, and the main *actor* described in a text. Both *topic* and *actor* can occur only once in each paragraph, as variables must be assigned to each paragraph mutually exclusive. As an example, the main issue discussed in a text about a protest could be the violence surrounding an event – either the violence against people or property. The main actor or rather actors could be the protesters if they are the ones described as resorting to violence. Or it could be the police or officials who try to stop the violence.[1] Taken together it becomes clear that *violence and protesters* or *violence and police* form two distinct *problem definitions*, although it is not yet clear who is responsible for the problem – that is what the other variables are for.

1.1. Topic

Topic means the main point that is discussed in a paragraph. If a paragraph contains more than one topic, the one that takes up more space is the main topic. If two or more are discussed equally, the one that is more specific is chosen. For example, most discussions about the spectacle during an event also include discussion of the event itself (see below for explanations of these two codes) but *spectacle* is arguably more specific. For ease of use, **codes are ordered from more specific to less specific within each variable.**

1.1.1. Clash

Confrontation with the police, not necessarily violent. If protesters are arrested, this usually triggers this code (except another one is more prevalent).

Examples:

- Officers in riot gear faced a hail of stones while a police helicopter filmed overhead as the demonstrators tried to break down fences around Hillgrove Farm near Witney. A police woman and a demonstrator were injured. Eight people were arrested for public order offences. (OWEN BOWCOTT (1997, November 17). Police and protesters clash at cat farm. The Guardian).

[1]It is also often described that the police started the violence. The example is just an illustration of a common combination of frame elements.

- "Then the officer beat Mr Tomlinson twice with his baton while he was on the ground. I remember thinking at that point, 'OK, that's an assault'. I didn't get a picture because an officer grabbed me by the arm" (Rayner, Gordon and Swaine, Jon (2009, April 09). RECONSTRUCTION; Final farewell before slow walk to death. The Daily Telegraph).
- Six people were arrested for disorderly conduct in east London last night outside a British National Party election rally in Tower Hamlets (Campbell, Duncan (1992, April 07). BNP ARRESTS. The Guardian).

1.1.2. Violence/Crime

Violence, vandalism and destruction of public or private property surrounding a protest. If police are involved, usually Clash is coded, except if police officers are targeted specifically (see 2nd example).

Examples:

- VIOLENCE broke out at the Prince of Wales's favourite hunt yesterday when 150 protesters clashed with riders near Malmesbury, Wiltshire. At least two members of the Beaufort Hunt and one demonstrator were injured, several cars were damaged and windscreens were smashed in the biggest confrontation for two years between those for and against field sports (NN (1996, February 18). Royal hunt clash. Sunday Times).
- The brutal attack on Police Constable Leslie Turner during Saturday's clashes with anti-fascist demonstrators added a grotesque twist to the difficulties already facing black police officers. Singled out by white protesters supposedly marching against racism, PC Turner was pelted with bricks on the ludicrous grounds that he was a "traitor" (NN (1993, October 19). The Thin Black Line. The Times).

1.1.3. Cause

The cause of why a protest took place or the goals of the protesters. A paragraph that falls into this code should try to answer the question why a protest took place. Usually this is done by interviewing one of the protesters or providing background information about the demands.

Examples:

- 'This is a protest against the super-rich with their own planes who are putting two fingers up to attempts by the rest of us who try to cut our carbon emissions by saying they will not takes off only continue to fly, but they will fly in the most carboninefficient way possible' (Charlotte Gill (2007, August 17). Airport protesters spreading their wings. Daily Mail).

- "Millions more are going to face poverty in retirement because ministers are walking away from promises they made. We are calling for a new, fairer system where everybody plays their part to provide a decent pension for everyone else." (Alison Purdy (2004, June 20). LONDON SEES BIGGEST-EVER PENSIONS MARCH; NO GOLDEN AGE CAMPAIGNERS WARN THAT MILLIONS FACE POVERTY. Independent on Sunday).

1.1.4. Protesters

Description of the protesters, especially their appearance, mental ability, visual deviance and oddities of the protesters (including pathologising their protest group, social movement or subculture) or their 'real' underlying motives (e.g., adolescent anger, lust for destruction, chaos and anarchy or astroturfing allegations).

Examples:

- Harris, who was bailed until next month by Highbury Corner magistrates, went on: 'My parents find the fact that I am a New Age traveller rather bizarre. I suppose that's not too surprising when you compare our lifestyles.' Harris's home is a tent, which he shares with his 23-year-old girlfriend Kendra Collier. A far cry from the plush detached family home (Tony Gallagher, David Connett, Tracey Harrison, Andrew Riley (1994, October 11). The ravers who called the tune. Daily Mail).
- Heather Gordon, 31, and her friend Jenny Myatt, 32, painted their faces with the EU's gold stars on a blue background and wrapped themselves in flags for the protest (Ben Quinn (2016, June 29). Crowds gather outside parliament to protest against Brexit. The Guardian).

1.1.5. Spectacle

Highlighting the entertaining or spectacle aspects (e.g., stunts or performances by protesters, eye-catching costumes and banners or the presence of celebrities).

Examples:

- Several celebrities, including Bianca Jagger, turned out to support the march (Not Author (2003, March 23). Terror fears as marchers jam the capital again. Mail on Sunday).

- The demonstrators walked from nearby Hatton Cross to Sipson - the village that will be lost if the new runway is built - to form a huge "NO" on a field visible to passing aircraft. Thousands of protesters gathered at the airport in West London to step up their long-running campaign (BY NICK OWENS (2008, June 01). RUNWAY? NO WAY SAYS CHAIN OF PROTEST. Sunday Mirror).

1.1.6. Protest tactics

Discussion about the tactics of protesters and why they were adopted. If there is a discussion if protest is the right tactic in general to accomplish a goal, this should be coded.

Examples:

- "Direct action is the only way to change opinion and change our attitudes to transport and car dependency before we suffer worse congestion, more asthma epidemics, and a complete reduction in our quality of life", said a spokesman yesterday (John Vidal (1995, August 05). FUMES GALORE AT ROAD PROTESTERS' BLOCKADE. The Guardian).
- "While protesters should be able to march peacefully to highlight their concerns, they should not be able to seriously disrupt the lives of Londoners and prevent them going about their daily business. (Sean O'Neill(2010, November 29). Think of consequences, police tell student protesters. The Times).
- Groups including Reclaim the Streets, the Anarchists' Federation and the Socialist Workers Party have called on protesters to plant seeds and bring tools for "guerrilla gardening" (Stewart Tendler (2000, April 29). Anarchists warned over May 1 protest. The Times).

1.1.7. Policing tactics

Discussion about how the police, security forces or laws should deal with protesters, including discussion about protest bans. When laws are discussed, the difference to judicial prosecution the focus is not on whether someone broke a specific law or is persecuted for it but whether it leads to suppression of protest or to reinstating public order.

Examples:

- But senior police chiefs say they need extra powers to act against the new type of anarchist action, which is arranged through the Internet and has no organisers who could be the target of any ban (Stephen Wright; Michael Clarke (2000, May 10). ANARCHISTS TO TARGET THE QUEEN. Daily Mail).

- “There are already tough laws in this area but we intend to strengthen these further to give the police additional powers to ensure that business and individuals can go about their lawful business without fear of violence or intimidation." (By Philip Johnston Home Affairs Editor (2001, January 18). Picketing laws to protect animal laboratory staff. The Daily Telegraph)
- There was not a riot officer in sight. One source said a van full of riot police had been on standby in Argyll Street, but had been stood down to avoid appearing heavy-handed (Jon Ungoed-Thomas; David Leppard; Jamie McGinnes; Rosie Kinchen; Lucy Osborne (2010, December 12). Police blunders put royals at risk. Sunday Times).

1.1.8. Response

Repeating or discussing the response of the target or third party to the protest demands.

Examples:

- The pharmaceuticals industry says animal experiments are legally and scientifically necessary before it begins testing potential new medicines on humans (STEPHEN FOLEY (2005, January 20). ANIMAL RIGHTS LOBBY DETERS MORE FIRMS. The Independent).
- A spokesman for Shell said: "Shell recognises certain organisations are opposed to our exploration programme Offshore Alaska, and we respect the right of individuals and organisations to engage in a free and frank exchange of views.” (Laurie Tuffrey (2012, July 17). Arctic protesters shut down Shell petrol stations. The Guardian).

1.1.9. Confrontation/Showdown

The protests are part of a confrontation between two groups (e.g., between political parties). Usually, a ‘showdown’ rhetoric is applied. Similar to a horse race rhetoric in reporting about elections, the reason for a protests is often moving into the background over the more newsworthy story of a fight between opponents.

Examples:

- As the flotilla towing the redundant platform headed through stormy seas, the controversy surrounding Shell's decision continued to escalate. While the British government again defended Shell's actions, opposition parties joined calls across Europe for a boycott of Shell petrol stations as a protest (MICHAEL CASSELL, Business Correspondent (1995, June 19). Shell pledges to go ahead with dumping oil rig as row grows. Financial Times).

- ANTI-roads protesters camped in the east Devon countryside are ready to do battle against eviction after contractors last week won the right to remove them from the route of a proposed dual carriageway on the A30, a main trunk road to the South-west (Geoffrey Gibbs (1996, October 14). A30 PROTESTERS DIG IN TO BATTLE THE BAILIFFS. The Guardian).

1.1.10. Public opinion

The protest represents a minority/majority of public opinion (operationalised through polls, interviews or reference to norms).

Examples:

- A Sunday Times poll last night showed only 33 per cent of Britons back an attack without United Nations support. The marchers, organised by the Stop the War Coalition, were largely a mix of pacifists, socialists and the pro-Palestinian lobby (Graham Johnson (2002, September 29). 350,000 SAY NO TO WAR; LONDON PROTEST BLASTS PM'S SADDAM PLAN. Sunday Mirror).
- A Mori poll last year found an overwhelming majority in favour of vivisection providing certain conditions were met (NN (2004, January 28). Animal wrongs. The Daily Telegraph).

1.1.11. Nuisance

The protest caused inconvenience to regular citizens and the government.

Examples:

- "It's a ruddy nonsense", said financial dealer James Wensley, on his way into the City. "Who the hell are these people? There's bad enough traffic anyway. What's the point of making it worse?" (John Vidal(1995, August 05). FUMES GALORE AT ROAD PROTESTERS' BLOCKADE. The Guardian).
- Outrage from thousands of would-be passengers at London City airport that nine young activists should delay them for an hour or two by rowing across a dock and occupying a taxiway. But consternation too, at the Daily Mail and even among some of the English liberal middle classes. (John Vidal (2016, September 07). The Black Lives Matter protesters were right: air pollution is a race issue. The Guardian).

1.1.12. Other activism

Other actions by the same group who initiated protest (e.g., letters or legal battles).

Examples:

- So far, more than 6,000 letters of protest have been sent, according to Nick MacKinnon, a mathematics teacher at Winchester College in Hampshire and organiser of the Campaign to Save Radio 4 Long Wave (Janine di Giovanni (1992, October 11). Angrey Radio 4 fans revolt over FM plan. Sunday Times).
- The equivalent of two million light bulbs were switched off at 9pm as householders answered UDM leader Roy Lynk's call for a protest blackout. 'It was noticeable but fairly small,' said the National Grid Company (NN (1992, October 19). From all walks of life; TORY STALWARTS MARCH IN STEP WITH THE MINERS. Daily Mail).

1.1.13. judicial prosecution

Prosecution of protest-related actions in court (e.g., court case about violence of protesters or police).

Examples:

- A PROTEST march strayed from its permitted route to avoid a riot, a court heard yesterday (Alan Erwin (2017, August 16). March 'left agreed route for safety of protesters'. Daily Mirror).
- Fourteen campaigners arrested in a dispute over tree-felling in Sheffield are to take legal action against South Yorkshire police (Josh Halliday (2017, March 22). Sheffield tree protesters to take legal action against police; Protesters detained for trying to stop contractors from chopping down trees to challenge legality of their arrest. The Guardian).

1.1.14. Effect of protest

Discussion about the effect a protest had or lack thereof (e.g., starting a public debate about a topic, leading to policy changes, etc.).

Examples:

- Dwarfed by monolithic banks, this small demonstration seemed impotent - but its target had been brought close to bankruptcy thanks to the extraordinarily effective campaign run by Shac (Cole Moreton (2001, January 21). FOCUS: ANIMAL RIGHTS: THE CAMPAIGNERS - 'BEFORE WE TARGET ANYWHERE, WE RESEARCH IT THOROUGHLY'. Independent on Sunday).

- AN EMERGENCY package of help for the livestock industry is being considered by the Government against a background of widening protests by British farmers angry at the impact of cheap beef imports on their crisis-torn industry (Philip Webster, and Michael Hornsby (1997, December 05). Help planned for farms hit by beef ban. The Times).

1.1.15. Event

Description of the event (e.g., size, marching route or what protesters did) but not highlighting the entertaining or spectacle aspects (see Topic Spectacle). This also includes how a protest is expected to develop (i.e. how many people will likely attend etc.).

Examples:

- Hundreds of mothers and children campaigning against cuts to maternity services marched through Westminster (NN (1998, March 12). Mothers' call. The Times).
- A CROWD of up to 70 people protested outside a hostel for the homeless after a report that a convicted paedophile had been housed there (NN (2000, September 01). Protest at hostel after newspaper's paedophile alert. The Times).

1.2. Actor

The actor in a paragraph should usually be easy to code. Look for a combination of a person or group and an activity. The activity might include protesting, fighting or talking. Again, if several actors are mentioned, the first one should be coded except if the second one takes up more space in the paragraph. Codes below only have an example when this was found helpful.

1.2.1. Protesters

The people engaging in a protest.

1.2.2. Police

Members of the police force engaging or clashing with protesters or commenting on a protest. Note that if policemen are protesting themselves, they are coded as protesters.

1.2.3. Officials

Representatives of government organisations.

1.2.4. Business

Representatives of business organisations.

1.2.5. Other Political Elite

Other political players such as members of the opposition party or regional parliaments.

Example: Sir David Naish, union president, said farmers demonstrating within the law to stop cheap imports of beef from Ireland and the continent were "supporting a most important cause" (Alison Maitland (1997, December 09). Farmers' union steps up pressure for aid in beef crisis. Financial Times).

1.2.6. Other

Other non-elite actors (e.g., motorists, local residents, counter-demonstrators or unidentified people).

Example: Martin Courtney and John Roe were among nine people who mounted the late-night attack to remove opponents of the M11 extension from a chestnut tree on a green in Wanstead, northeast London (Martin Courtney and John Roe (1994, September 06). Two jailed for fire attack on protesters. The Times).

2. Causal Attribution

Causal attribution and ***Moral evaluation*** – the second and third frame elements – are closely linked to each other. Causal attribution is conceptualised to contain an attribution of who is responsible for a described *benefit* or a *risk* or both. Moral evaluation entails said benefits and risks. Therefore, it usually makes sense to code ***Moral evaluation*** before ***Causal attribution***. It is, however, also possible that a paragraph does contain a causal attribution without a moral evaluation, if a risk or benefit have been described earlier in the article and the paragraph specifically blames someone.

In above's example, a text could portray protesters as being responsible for the risk that reported violence caused. They do not necessarily need to be the main actor in the text for that. For example, the risk could be that public safety was endangered or property was destroyed. The police could be portrayed as the actor and as being responsible for the benefit of ending the violence by detaining violent protesters. However, a news report could also flip the attribution roles: risking public safety could be attributed to the police if they started to clash with protesters unprovoked and without necessity. The protesters could be portrayed to cause a benefit, for example, if their message is portrayed as important and they are seen as demonstrating for a good cause. Depending on the benefit and risk and who is deemed responsible for that, the example above – *violence* being the topic and *protesters* being the actor – can be modified substantially: protesters might be the main actor in a story about a peaceful protest that turned into violence, yet the risk of endangering public safety might be attributed to few overambitious police officer. Again, there can only be one *risk* or *benefit* and *risk/benefit attribution* within a paragraph. Yet an **article** might discuss several *risks/benefits* or make different actors responsible for the same *risk*/*benefit* **over several paragraphs**.

2.1. Benefit Attribution

2.1.1. Police

Police is responsible for the benefit (e.g., by reinstating order).

Examples:

- More than 170 supporters of the far-right English Defence League were arrested in Westminster yesterday as police moved to prevent a repeat of the violent clashes that took place on Remembrance Day last year (KEVIN RAWLINSON (2011, November

12). Police foil EDL plan to target St Paul's; 172 arrested as far-right group gathers in Whitehall to march on Occupy City protesters. The Independent).

- But Tory MP Peter Bottomley said: "Officers were fully justified in making sure the crushing stopped. I believe they saved many serious casualties, if not fatalities." (DON MACKAY AND NATHAN YATES (2004, September 16). TOFF WITH THEIR HEADS: BAN FURY: TALLY HOOLIGANS; POLICE HIT AS HUNT YOBS HURL FIREWORKS IN RIOT. Daily Mirror).

2.1.2. Protesters

Protesters are responsible for a benefit (e.g., by being entertaining or initiating debate).

Examples:

- Mr Cogswell said he welcomed the media focus the march would bring because he wanted people to debate the issues and bring the campaign to the world's attention. (Rosie Murray-West (2006, February 21). Showdown in Oxford as students face opponents of animal tests. The Daily Telegraph).
- A member of the Fairy Army said that he was there to have fun and to "spread a little magic" (David Lister and Shirley English (2005, July 05). Anarchists accuse the police of overreacting to parade of 'fun'. The Times).

2.1.3. Officials

Officials are responsible for a benefit (e.g., by preventing chaos protesters try to inflict).

(Note: almost never used).

2.2. Risk Attribution

2.2.1. Protesters

Protesters are responsible for risk (e.g., vandalism, attacks on police or nuisance).

Examples:

- Singled out by white protesters supposedly marching against racism, PC Turner was pelted with bricks on the ludicrous grounds that he was a "traitor" (NN (1993, October 19). The Thin Black Line. The Times).
- But one officer said: "They [protesters] were tying fireworks together, lighting them and hurling them in our faces." (DON MACKAY AND NATHAN YATES (2004,

September 16). TOFF WITH THEIR HEADS: BAN FURY: TALLY HOOLIGANS; POLICE HIT AS HUNT YOBS HURL FIREWORKS IN RIOT. Daily Mirror).

2.2.2. Police

Police are responsible for risk (e.g., unnecessary clashes with peaceful protesters).

Examples:

- Police baton-charged the crowd, leaving some demonstrators dazed and bleeding and leading to accusations of heavy-handedness and brutality. (DON MACKAY AND NATHAN YATES (2004, September 16). TOFF WITH THEIR HEADS: BAN FURY: TALLY HOOLIGANS; POLICE HIT AS HUNT YOBS HURL FIREWORKS IN RIOT. Daily Mirror).

2.2.3. Business

Business actors are responsible for risk.

Examples:

- We strike up an incongruous conversation with a man in his sixties, a Telegraph subscriber, who has come to express his support. "I thought these people in the City were my friends," he says. "But what they've been doing for the last 15 years is downright greedy." (Iain Hollingshead (2011, October 20). YOU WANT WHAT, EXACTLY?; The St Paul's tent villagers know what's wrong with the world - but have they outstayed their welcome, asks Iain Hollingshead. The Daily Telegraph).

2.2.4. Officials

Officials are responsible for risk (e.g., by signing a bad law).

Examples:

- Andy, a regular at Fairmile, gave reasons for his protest: "Primarily it's to protect the land. Secondly it's to make the Government rethink their earth-raping road-building programme (Geoffrey Gibbs (1996, October 14). A30 PROTESTERS DIG IN TO BATTLE THE BAILIFFS. The Guardian).

2.2.5. Media

The media are responsible for risk (e.g., problematic reporting).

Examples:

- "There is an absolute legitimate concern that the BBC is giving undue priority to those people who are protesting about cuts and very limited coverage of exactly why the cuts are necessary. (Tom Whitehead (2011, March 28). BBC accused of bias in row over coverage. The Daily Telegraph).

2.2.6. Other Pol. Elite

Other political players such as members of the opposition party or regional parliaments are responsible for a risk (almost never coded).

2.2.7. Other

Other actors are responsible for a risk, often by attacking protesters (e.g., motorists, local residents, counter-demonstrators or unidentified people). This code might be necessary if an actor is important in the paragraph but does not fit in any of the categories. An example:

“Martin Courtney and John Roe were among nine people who mounted the late-night attack to remove opponents of the M11 extension from a chestnut tree on a green in Wanstead, northeast London” (Martin Courtney and John Roe (1994, September 06). Two jailed for fire attack on protesters. The Times).

The two mentioned actors fall in none of the categories. They were hired thugs who were paid to intimidate protesters. In practice this code was used rarely.

3. Moral Evaluation

Moral Evaluation describes, as was described above, the risk or benefit mentioned in a paragraph. Meant here is the risk or benefit caused or denounced during a protest.

3.1. Benefit

3.1.1. Reinstating public order

Usually attributed to police who reinstate public order after it was strained by the protest.

Examples:

- More than 170 supporters of the far-right English Defence League were arrested in Westminster yesterday as police moved to prevent a repeat of the violent clashes that took place on Remembrance Day last year (KEVIN RAWLINSON (2011, November 12). Police foil EDL plan to target St Paul's; 172 arrested as far-right group gathers in Whitehall to march on Occupy City protesters. The Independent).

3.1.2. (Just) Cause

Usually attributed to protesters when their actions are seen as part of a struggle for a good cause.

Examples:

- "It sends a clear message to the Home Office that they are not going to get away with their inhumane barbaric asylum policies in Scotland." (GARY ANDERSON (2005, November 03). MULLAN IN DAWN RAID ON ASYLUM CENTRE; ACTOR JOINS PROTEST SWOOP ON IMMIGRATION HQ. Daily Mirror).
- "Direct action is the only way to change opinion and change our attitudes to transport and car dependency before we suffer worse congestion, more asthma epidemics, and a complete reduction in our quality of life", said a spokesman yesterday. (John Vidal (1995, August 05). FUMES GALORE AT ROAD PROTESTERS' BLOCKADE. The Guardian).

3.1.3. Initiated public debate

Usually attributed to protesters when their actions have caused a (necessary) public debate.

Examples:

- "Absolutely," she said. "No reservations at all." At last the public was beginning to question road schemes which entailed ripping through the countryside, she added. (Louise Jury and Clare Garner (1997, January 29). Animal pulled out of labyrinth; My A-levels can come later, says Devon roads protester, aged 16. The Independent).

3.2. Risk

Again, if there are several risks, the one that is more prevalent in a paragraph should be coded. If two or more risks appear equally important, the more specific one should be chosen. The risk *breaking laws*, for example, was deemed less specific than property destruction, which in turn is less specific than public safety.

3.2.1. Grievance

The grievance which the protest is addressing is the risk. Meant is that the risk described in a paragraph is the (potential) alleged negative effect of a policy or practice that is the focus of a protest. The code is usually assigned together with the Topic Cause, if a (perceived) negative effect of the contested policy or practice is mentioned.

Examples:

- Andy, a regular at Fairmile, gave reasons for his protest: "Primarily it's to protect the land. Secondly it's to make the Government rethink their earth-raping road-building programme. (Geoffrey Gibbs (1996, October 14). A30 PROTESTERS DIG IN TO BATTLE THE BAILIFFS;. The Guardian)
- Why is this government so intent on killing off every institution we have? Small post offices are a lifeline to the elderly who cannot walk the distance to a main post office, and who often don't have cars, or cannot drive because of failing eyesight. (Sally Emerson (2008, March 04). Why I've become a placard waver!; FROM POST OFFICES TO PARLIAMENT, THE MIDDLE CLASSES ARE IN REVOLT. Daily Mail).

3.2.2. Public safety

The protest as a risk for public safety.

Examples:

- The men, who have lived in Britain since 1989, were punched by police and one was hit with a truncheon after they were arrested for violent disorder during a protest outside a community centre in north London. (Richard Ford (1996, June 14). Kurds win Pounds 150,000 for false arrest. The Times).

- 3.25PM: About 75 protesters reach the roof top of the Tower. A fire extinguisher is thrown at cops. (NN (2010, November 11). Timeline of havoc. The Sun).

3.2.3. Property destruction

Property was destroyed due to protest.

Examples:

- More than 50 were arrested after officers in riot gear beat back a group who short-circuited an electric fence in a bid to break into the site. (GRANT HODGSON(2008, August 10). ECO WARRIORS; RIOT POLICE BEAT OFF CROWD STORMING POWER STATION. Sunday Mirror).
- Parliament Square was overrun, with people urinating on Churchill''s statue, spraying walls and burning benches. This was before the votes had even been counted. (Sean O'Neill, Valentine Low (2010, December 11). Scare raises urgent security questions only months before the royal wedding; London riots. The Times).

3.2.4. Harassment

Protesters are harassing people instead of engaging in arguments (e.g., screaming at clients in family planning facilities instead of campaigning for anti-abortion laws or intimidating workers of animal testing facilities instead of campaigning for stricter animal welfare laws).

Examples:

- His announcement was part of a concerted government effort to confront the protesters. Earlier this week, Downing Street condemned the "intimidation, violence and thuggery" being used against HLS. (Philip Johnston (2001, January 18). Picketing laws to protect animal laboratory staff. The Daily Telegraph).
- Anti-abortion campaigners have been known to confront women outside family planning facilities with pictures of foetuses and to host "abortion vigils" where they pray for people to change their minds on seeking a termination. Following pressure from more than 110 cross-party MPs to bolster protections for vulnerable women, the Home Secretary has vowed to consider new police-and-civil powers to clamp down on aggressive protesters. (LIZZY BUCHAN (2017, November 27). Home Secretary considers new powers to protect women seeking abortions. The Independent).

3.2.5. Breaking laws

Breaking the law but neither destroying anything nor harming anyone (e.g., entering parliament without permission).

Examples:

- Scotland Yard said seven men went over the wall but two were arrested before they could make the climb. Both face breach-of-the-peace charges. (DAVID OWEN (1994, November 05). Commons rooftop protesters try to stay in touch. Financial Times).
- However, the number of "home visits" - where protesters gather outside the home of a company director or employee - declined last year and there were at least 124 arrests, almost three times as many as in the previous year. For the first time, there were no physical attacks on workers associated with vivisection. (STEPHEN FOLEY (2005, January 20). ANIMAL RIGHTS LOBBY DETERS MORE FIRMS. The Independent).

3.2.6. Decay of morals or other social norms

The protesters are a fringe group of freaks who set a bad example for others and disturb the general political consensus. A variation which is subsumed under this code is what Di Cicco (2010) calls “strict father morality, in which obedience to legitimate authority is moral behavior and disobedience is both immoral and threatening to the social order.” (p. 136).

Examples:

- Parliament Square is in danger of becoming a shanty town that is almost impossible to police, London's deputy mayor for policing told The Times as Tamil protesters planned to defy a deadline to remove an encampment. (Sean O'Neill (2009, May 25). 'Shanty town' fears as Tamils defy their deadline to move on. The Times).
- London and Glasgow will echo to the slogans of the morally deluded and the self-consciously caring. Papoose-wearers, manic recyclers, the priggish, the cranky, nudists and Woodcraft Folk will march this Saturday in a cloud of outrage. Peaceniks, marshalled by the Stop the War coalition, claim to march for the majority. They do not. (Robbie Millen (2003, February 12). March for anything, anywhere. The Times).

3.2.7. Nuisance

Protest is bothersome to citizens and the government (even if it may still be regarded as a legitimate practice overall).

Examples:

- "It's a ruddy nonsense", said financial dealer James Wensley, on his way into the City. "Who the hell are these people? There's bad enough traffic anyway. What's the point of making it worse?” (John Vidal (1995, August 05). FUMES GALORE AT ROAD PROTESTERS' BLOCKADE. The Guardian).
- ST PAUL'S Cathedral will remain closed for the foreseeable future because of the anti-capitalist protest camp on its doorstep. Activists refused to budge yesterday - with many pledging to remain until CHRISTMAS and beyond. It came as thousands of Sunday worshippers were turned away. The London landmark shut its doors on Friday for the first time since the Second World War. (NN (2011, October 24). St Paul's stays shut as protest continues. The Sun).
- PETER JACKSON, a farmer, revved the engine of his big tractor as he inched past irate motorists in a long traffic jam caused by fellow fuel protestors. He said: "I think we're doing the British public a service." (Cahal Milmo (2000, November 11). DEFIANCE THAT SLOWLY MELTED ON THE CONVOY'S LONG ROAD SOUTH. The Independent).

3.2.8. Harm discussion/cause

The protest or actions of protesters harm the discussion about a topic (e.g., by shouting down opponents or by distracting from the cause). Or they are hurting the actual cause by causing an adverse effect.

Examples:

- Unfortunately, the hysteria surrounding the Dungavel protests is hampering such a sensible debate on asylum and immigration. (GEORGE FOULKES (2003, September 16). DUNGAVEL THE TRUTH. Daily Mail).
- "I would have been happy to talk to the students afterwards, but then the smoke bomb was set off and that ended the conversation." (Andrew Hough (2011, June 08). AC Grayling forced to flee smoke bomb protest at Foyles debate. The Daily Telegraph).

3.2.9. Trivializing (political) discussion

The protesters spoil serious (political) discussions with their childish, insane or uninformed arguments or false claims. Similar to harm discussion but not as strong. Instead of making the discussion impossible for everyone they only allegedly waste everyone’s time.

Examples:

- They wanted more government support but less government interference; to be left alone but not to be ignored. Yesterday's Countryside Alliance march may not have been the most coherent of political protests but there could be no doubt about its scale or the passion expressed by some of the 400,000 protesters. (MARIANNE BRUN-ROVET and JOHN MASON (2002, September 23). Countryside protesters enjoy field day in city. Financial Times).
- One man, wearing the mask of the Anonymous UK group, best known for its hacking activities, tells me that the world's banks are owned by five or six people, "the Rothschilds or something". I lose count of the people who want some form of ill-defined, bottom-up socialism, without being able to articulate what this might look like in practice. It is hard not to feel that this is a generalised squeal of pain about the unfairness of life, rather than a campaign with an achievable goal. (Iain Hollingshead (2011, October 20). YOU WANT WHAT, EXACTLY? The Daily Telegraph).

3.2.10. Bad for business

Protests have a negative impact on business revenues.

Examples:

- He said: 'This had had a very serious effect on us. Despite the petrol companies dropping five pence a litre off the retail price, we have been hit badly by the drop in fuel sales and the knock-on effect on shop sales.' The Scottish experience was mirrored throughout the UK. (Ian Smith (2000, August 02). Drivers back petrol boycott to fuel revolt against prices. Daily Mail).
- The city centre was brought to a standstill and businesses on one of Britain's most famous shopping streets closed their doors with their staff locked inside. (David Lister and Shirley English (2005, July 05). Anarchists accuse the police of overreacting to parade of 'fun'. The Times).

3.2.11. Costs of demonstrations

The costs of a demonstration (e.g., clear up and police deployment costs of demonstrations) including what this might cause (e.g., burden public budget or spread police thin for other tasks).

Examples:

- The police and court costs of the operation since the protest began in September have been put at Pounds 1 million. Chief Supt Stuart Giblin said: "We have spent, at a conservative estimate, at least Pounds 150,000 today. That is money which could have been put to much better use. (Andrew Pierce (1994, February 17). Army of bailiffs beats M11 squatters. The Times).

- Animal rights protests at Huntingdon Life Sciences have cost Cambridgeshire Constabulary Pounds 3.8 million and left them struggling to cover normal policing (Stewart Tendler (2003, March 13). Lab protest cost police Pounds 3.8 million. The Times).

3.2.12. Suppression/Censorship

Usually attributed to police or officials who allegedly try to undermine protest or make it impossible.

Examples:

- A spokesman for the organisers said: "The new legislation will severely restrict the ability to organise effective and spontaneous public demonstrations." (NN (2010, May 11). UNIONS IN WARNING ON PARADES LAW. Daily Mirror).
- On Wednesday, senior judges heard an appeal that Gloucestershire police behaved illegally when they stopped three coachloads of protesters travelling to a US military base to protest against the Iraq war. (John O'Farrell (2004, December 10). Comment & Analysis: On the road to Fairford. The Guardian).
- The Labour Party has also tried to gag the marchers by banning the use of a PA system outside the SECC. (JUSTINE SMITH (2003, February 15). THE WORLD AGAINST THE WAR: FEB 15, 2003: A MILLION MARCHING; PEACE DEMOS ARE BIGGEST SINCE VE-DAY. Daily Mirror).

4. Treatment

Treatment is conceptualised to contain either positive or negative ***judgement***. In the case of protests, this usually means that if a protest is judged positively, the treatment recommendation is that protests should continue until a stated goal is reached. The treatment recommendation might be explicit, but usually it is not. Rather, if judgement is negative, the treatment recommendation is that protesters should disperse and go home, or that other actors should act tougher to bring protests to an end. Speaking from experience, *judgement* was rarely assigned in coding. When it comes to protest in the UK, journalists tend to agree that explicitly arguing for a protest to stop would be out of line. This does not mean though that it is not often implied by highlighting the risks and being silent about any advantages. A frame might thus be judgmental of protest even without explicitly stating that opinion, which is captured through occurrence of variables in the other frame elements. Therefore, implicit judgement should not be coded.

4.1. Judgement

4.1.1. Negative

The protests were/are bad (e.g., protesters should go home or someone should stop them).

Examples:

- He said he may consider asking Home Secretary Jack Straw to ban any demonstrations which might result in 'major public disorder' a move which would infuriate civil liberties groups. (Stephen Wright; Michael Clarke (2000, May 10). ANARCHISTS TO TARGET THE QUEEN. Daily Mail)
- I don't know much about ballet, but I know what I don't like. I don't like the British National Party. But I cannot stand the campaign to have Simone Clarke sacked from the English National Ballet for supporting the BNP. (Mick Hume (2007, January 16). Hounding the BNP ballerina is tutu absurd. The Times)
- "While protesters should be able to march peacefully to highlight their concerns, they should not be able to seriously disrupt the lives of Londoners and prevent them going about their daily business." (Sean O'Neill (2010, November 29). Think of consequences, police tell student protesters. The Times)

4.1.2. Positive

The protests were/are good (e.g., protests made aware of a problem).

Examples:

- "They are good, law-abiding lads - all they want to do is make an honest living and feed the kids like everyone else," he volunteered in the wheelhouse as the crackling port-control radio broadcast the threat of legal action. (PETER HETHERINGTON (1993, March 23). COURT ENDS FISHING PROTEST. The Guardian)
- Terry Deakin, co-owner of Advance Coils in Warrington, said the problems had already cost him up to pounds 5,000 in lost business, but he still maintained the protesters had done a "wonderful job. We have been in support of the drivers," he said. "It was short, sharp and got the message home, but they were right to call it off when they did. It would have hurt if it had carried on." (Rob Griffin, Jamie Wilson and Angelique Chrisafis (2000, September 16). Fuel crisis: Cost of dispute could top pounds 1bn, say firms: Protests called off 'in the nick of time' to avoid factory closures and the laying off of workers. The Guardian)

B Additional Information on the Cleaning of the News Media Database

Table B.1: Topic Model Used for Cleaning

	topic 1	topic 2	topic 3	topic 4	topic 5	topic 6
Name	Literature & Film	Ireland	Sport	War Europe	Women's Rights	Miscella-neous
Relevant	no	yes	no	no	yes	no
1	one	irish	england	german	women	say
2	even	ireland	rugbi	germani	men	peopl
3	seem	dublin	play	war	sex	can
4	like	said	player	italian	woman	go
5	much	last	game	itali	sexual	think
6	man	year	team	berlin	girl	get
7	yet	protest	wale	europ	femal	don
8	might	euro	tri	nazi	male	like
9	never	murphi	cup	year	abort	want
10	though	cork	coach	east	rape	just

Table B.1: Topic Model Used for Cleaning (continued)

	topic 1	topic 2	topic 3	topic 4	topic 5	topic 6
Name	Literature & Film	Ireland	Sport	War Europe	Women's Rights	Miscella-neous
Relevant	no	yes	no	no	yes	no
1	one	irish	england	german	women	say
2	even	ireland	rugbi	germani	men	peopl
3	seem	dublin	play	war	sex	can
4	like	said	player	italian	woman	go
5	much	last	game	itali	sexual	think
6	man	year	team	berlin	girl	get
7	yet	protest	wale	europ	femal	don
8	might	euro	tri	nazi	male	like
9	never	murphi	cup	year	abort	want
10	though	cork	coach	east	rape	just

Table B.1: Topic Model Used for Cleaning (continued)

	topic 7	topic 8	topic 9	topic 10	topic 11	topic 12
Name	Movies	Business	Judicial	Sport	Protest (Infrastructure)	Science
Relevant	no	no	no	no	yes	no
1	film	compani	law	race	airport	research
2	star	market	court	year	transport	scienc
3	play	busi	right	hors	train	use
4	movi	year	legal	win	london	univers
5	actor	sale	case	last	passeng	scientist
6	show	group	rule	first	flight	human
7	pm	share	act	two	travel	professor
8	seri	profit	govern	one	rail	studi
9	charact	price	judg	play	airlin	test
10	director	industri	justic	world	road	develop

Table B.1: Topic Model Used for Cleaning (continued)

	topic 13	topic 14	topic 15	topic 16	topic 17	topic 18
Name	Politics/Protests Foreign	Sport	Church & Royalty	Financial Crisis	Fuel Protest	History and Obituaries
Relevant	no	no	no	no	yes	no
1	govern	minut	church	bank	per	year
2	elect	goal	queen	market	cent	becam
3	presid	ball	princ	econom	price	first
4	polit	half	royal	economi	year	later
5	parti	first	king	rate	said	die
6	countri	second	st	financi	increas	work
7	minist	game	christian	year	rise	born
8	power	time	bishop	debt	fuel	time
9	year	score	charl	growth	last	age
10	opposit	refere	cathol	crisi	oil	one

Table B.1: Topic Model Used for Cleaning (continued)

	topic 19	topic 20	topic 21	topic 22	topic 23
Name	Gossip	EU	Family	Property	Politics
Relevant	no	no	no (but Fathers 4 Justice)	yes	yes
1	wear	european	children	hous	said
2	fashion	french	famili	citi	plan
3	look	eu	mother	build	govern
4	dress	franc	year	home	yesterday
5	design	europ	father	london	decis
6	cloth	britain	parent	new	meet
7	like	british	life	local	last
8	new	countri	live	street	propos
9	show	minist	child	plan	minist
10	shop	pari	home	properti	protest

Table B.1: Topic Model Used for Cleaning (continued)

	topic 24	topic 25	topic 26	topic 27	topic 28	topic 29
Name	Finance	Riot	Business	Scandal	Literature	Crime
Relevant	no	yes	no	no	no	yes
1	pound	polic	work	report	book	prison
2	money	said	busi	said	write	court
3	million	offic	say	investig	publish	year
4	pay	protest	manag	alleg	read	murder
5	year	attack	compani	claim	novel	sentenc
6	cost	riot	servic	inquiri	writer	jail
7	bank	fire	peopl	public	stori	case
8	fund	violenc	need	evid	author	trial
9	paid	peopl	organis	case	life	said
10	charg	night	job	inform	english	charg

Table B.1: Topic Model Used for Cleaning (continued)

	topic 30	topic 31	topic 32	topic 33	topic 34	topic 35
Name	Political Conflict	Local Politics	Media	Britain	Protest	Bank
Relevant	no	no	yes	yes	yes	no
1	mr	council	bbc	polit	protest	execut
2	said	local	news	peopl	said	compani
3	yesterday	west	televis	social	peopl	bank
4	ms	north	radio	world	demonstr	sharehold
5	ad	london	programm	new	group	chief
6	protest	town	broadcast	power	polic	pay
7	last	south	show	cultur	london	year
8	told	east	media	nation	campaign	director
9	night	john	tv	chang	march	share
10	claim	counti	channel	societi	organis	investor

Table B.1: Topic Model Used for Cleaning (continued)

	topic 36	topic 37	topic 38	topic 39	topic 40
Name	Parks	Cars	Miscellaneous	Africa	Blair/Brown
Relevant	yes	no	no	no	no
1	garden	car	can	world	blair
2	tree	drive	one	africa	minist
3	land	driver	may	countri	mr
4	villag	road	even	south	brown
5	year	vehicl	much	african	prime
6	plant	engin	might	aid	labour
7	site	speed	time	intern	toni
8	road	motor	make	govern	govern
9	local	one	seem	peopl	gordon
10	green	new	way	nation	polit

Table B.1: Topic Model Used for Cleaning (continued)

	topic 41	topic 42	topic 43	topic 44	topic 45	topic 46
Name	Sarcastic Protest Reports	Fox Hunting & Fishing Protests	Sport	War	Israel	Politics - Fiscal
Relevant	yes	yes	no	no	yes	yes
1	like	hunt	player	war	israel	tax
2	get	fish	team	iraq	isra	govern
3	one	fox	england	russia	palestinian	cut
4	can	ban	footbal	russian	jewish	pension
5	say	countrysid	unit	us	jew	year
6	just	anim	unit	presid	peac	pay
7	go	year	chelsea	saddam	arab	public
8	don	bird	cup	putin	gaza	budget
9	re	rural	play	un	east	spend
10	think	countri	arsenal	moscow	west	benefit

Table B.1: Topic Model Used for Cleaning (continued)

	topic 47	topic 48	topic 49	topic 50	topic 51	topic 52
Name	Environ-ment	Anecdotes	Union	US Politics	Immigra-tion	Food
Relevant	yes	yes	no	yes	yes	yes
1	energi	said	union	american	polic	food
2	power	one	worker	us	offic	cook
3	gas	day	strike	presid	home	eat
4	oil	time	work	america	secur	restaur
5	nuclear	go	said	trump	said	wine
6	water	told	pay	new	asylum	drink
7	climat	just	action	state	crime	make
8	environ-ment	back	job	york	forc	chef
9	compani	ask	staff	bush	use	can
10	industri	went	industri	clinton	immigr	tast

Table B.1: Topic Model Used for Cleaning (continued)

	topic 53	topic 54	topic 55	topic 56	topic 57
Name	Terror	Tories	Internet	Opposition Parties	Asia
Relevant	no	yes	no	no	no.yes
1	muslim	minist	use	parti	china
2	islam	tori	can	labour	chines
3	terror	mps	internet	elect	kong
4	attack	lord	comput	vote	hong
5	pakistan	labour	technolog	tori	beij
6	india	govern	onlin	polit	japan
7	terrorist	common	phone	conserv	japanes
8	al	mp	system	leader	year
9	british	secretari	mobil	voter	countri
10	indian	parti	new	campaign	foreign

Table B.1: Topic Model Used for Cleaning (continued)

	topic 58	topic 59	topic 60	topic 61	topic 62	topic 63
Name	Education	Middle East	Miscella-neous	Minorities	Healthcare	Northern Ireland
Relevant	no	yes	no	yes	yes	yes
1	school	iran	one	black	health	ireland
2	student	al	like	peopl	hospit	northern
3	univers	regim	back	white	doctor	ira
4	educ	syria	look	gay	patient	belfast
5	teacher	govern	day	right	nhs	sinn
6	year	saudi	hand	group	medic	fein
7	children	countri	can	communiti	drug	unionist
8	colleg	arab	head	racist	care	protest
9	pupil	iranian	around	anti	year	peac
10	parent	foreign	time	racism	treatment	republican

Table B.1: Topic Model Used for Cleaning (continued)

	topic 64	topic 65	topic 66	topic 67	topic 68	topic 69
Name	Music	Travel	Art	Miscellaneous	Sport	Military
Relevant	no	no	no	no	no	yes
1	music	hotel	art	peopl	england	war
2	song	island	work	can	cricket	british
3	band	night	artist	even	test	militari
4	record	day	paint	now	australia	soldier
5	album	includ	exhibit	one	match	forc
6	rock	holiday	museum	us	ball	armi
7	year	travel	galleri	mani	play	defenc
8	pop	beach	show	must	first	troop
9	new	two	design	know	day	arm
10	singer	tour	photograph	like	team	command

Table B.1: Topic Model Used for Cleaning (continued)

	topic 70	topic 71	topic 72	topic 73	topic 74	topic 75
Name	Animal Right Protest	Events	Football	Scotland	Sport	Theatre
Relevant	yes	no	no	yes	no	no
1	anim	uk	club	scotland	olymp	play
2	farmer	www	footbal	scottish	sport	theatr
3	farm	pm	fan	glasgow	game	music
4	dog	pound	leagu	edinburgh	world	perform
5	protest	co	player	celtic	team	opera
6	said	festiv	manag	ranger	year	stage
7	right	day	game	snp	athlet	danc
8	campaign	free	season	scot	race	work
9	agricultur	children	unit	first	gold	product
10	food	com	support	heart	medal	london

Table B.2: Replaced Phrases

a clear demonstration	he still protests his innocence	protestations of
a clear demonstration of the	her protest	protestations of innocence
a demonstration of	her protests	protestations that
a formal protest	his protest	protestations to
a protest letter	his protestations	protestations to the
a protest vote	his protestations of	protested his
a Protestant	his protests	protested his innocence
after protesting	howls of protest	protested that

Table B.2: Replaced Phrases *(continued)*

after protests from	howls of protest from the	protested their innocence
ago in protest	I protest	protesting her innocence
and Protestant	in demonstrating	protesting his
and Protestants	in jail protesting his innocence	protesting his innocence
angry protests from	in protest	protesting that
another demonstration of	In protest	protesting their
any protest	in protest after	protesting their innocence
believed his protestations of innocence	in protest against the	protests from
by demonstrating	in protest and	protests his
by demonstrating that	in protest at	protests his innocence
Catholic and Protestant	in protest at the	protests last year
Catholics and Protestants	in protest over	provoked protests from
clear demonstration	is a demonstration	public protest
clear demonstration of	is demonstrating	quit in protest
continue to protest my innocence	is in protest	quit in protest at the
continue to protest their innocence	is still protesting his innocence	raised in protest
continued to protest her innocence	left in protest	resign in protest
continued to protest his innocence	letter of protest	resigned in protest
continued to protest their innocence	letter of protest to the	resigned in protest at the
continues to protest her innocence	letters of protest	resigning in protest
continues to protest his innocence	month in protest	right to protest
continuing to protest his innocence	not protest	right to protest , but
demonstrate a	of demonstrating	she protests
demonstrate how	of demonstrating that	sparked protests
demonstrate that	of protest	still protesting his
demonstrate that the	of protest and	still protesting his innocence

Table B.2: Replaced Phrases *(continued)*

demonstrated that	of protest at	storm of protest
demonstrated the	of protest from	strike to protest his innocence
demonstrates that	of protest in	strong protests
demonstrating a	of protest over	tax in protest
demonstrating his	of protest to	the Protestant
demonstrating how	of Protestant	the Protestant community
demonstrating how to	of Protestants	the Protestants
demonstrating its	of protests from	the protestations of
demonstrating that	of the Protestant	the protests of
demonstrating the	official protest	Tiananmen Square protests
demonstrating their	out in protest at the	to demonstrate his
demonstration of the	party in protest	to demonstrate how
demonstration plant	prompted protests	to demonstrate that
demonstration project	protest at their	to demonstrate the
demonstration projects	protest her innocence	to demonstrate their
demonstrations of	protest his	to protest his
Despite his protestations of innocence	protest his innocence	to protest their
despite protests	protest letter	to protest to
Despite protests	protest letters	to protests from
despite protests from	protest song	under protest
Despite protests from	protest songs	was in protest
doth protest too	protest that	was still protesting his innocence
fierce protests	protest their	wave of protest
following protests from	protest their innocence	way of demonstrating
formal protest	protest too much	week in protest
furious protests from	protest vote	who protests his innocence
government in protest	Protestant	who still protests his innocence
has always protested his innocence	Protestant and	will demonstrate
has been protesting his innocence	Protestant community	will vigorously protest his innocence

Table B.2: Replaced Phrases *(continued)*

has protested	Protestants	with protest
have always protested their innocence	Protestants and	without protest
he protested	Protestants and Catholics	would protest
he protests	protestations about	year in protest
He protests his innocence	protestations by	yesterday in protest

Table B.3: Phrases Used to Remove Articles

a dirty protest	cooking demonstrations	pro-democracy protesters
and cookery demonstrations	dirty protest	pro-democracy protesters in
cookery demonstration	pro-democracy demonstrations	pro-democracy protests
cookery demonstrations	pro-democracy demonstrators	pro-democracy protests in

C Regression Results for Change of Frame Share Over Time

	Delegitimising				Legitimising		
	Law & Order	Troublemakers	Decay of Morals	Nuisance	Cause & Grievances	Righteous Struggle	Police Violence
(Intercept)	0.080***	0.262***	0.341***	0.195***	0.256***	0.109***	0.231***
	(0.008)	(0.016)	(0.015)	(0.010)	(0.019)	(0.007)	(0.032)
year	−0001	0.001	0.000	−0000	0.006***	0.002***	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
R^2	0.061	0.023	0.003	0.000	0.437	0.428	0.016
Adj. R^2	0.022	−0018	−0038	−0042	0.414	0.404	−0026
Num. obs.	26	26	26	26	26	26	26

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{+}p < 0.1$

Table C.1: Regression Results for Change of Frame Share Over Time

D Recoded Protest Event Data

Table D.1: Employed Subset of the MMP Data by Clark and Regan (2019) with Additional and Recoded Variables

id	date	date_end	keywords	ideology
2001992001	1992-02-12	1992-02-12	solicitors, Law Society	right
2001992002	1992-02-12	1992-02-12	Students, higher education, government cuts	left
2001992003	1992-02-13	1992-02-13	British Field Sports Society, Master of Foxhounds Association, Piccadilly Hunt, field sports, fox hunting	right
2001992005	1992-03-01	1992-03-01	Hampshire, Twyford Down, Friends of the Earth	left
2001992006	1992-03-31	1992-03-31	Belfast, schoolchildren, Grosvenor High	neither
2001992007	1992-04-06	1992-04-06	PARENTS, education, schools, educational	left
2001992008	1992-05-31	1992-05-31	Bomber Harris, RAF, St Clement Danes	left
2001992009	1992-07-07	1992-07-07	fishermen, Sea Fish Conservation Bill	right
2001992010	1992-07-16	1992-07-17	Bristol, Hartcliffe estate	left
2001992012	1992-08-21	1992-08-21	Middlesex, nurses, Hospital, patients, Nupe, health union, Cohse, Nalgo	left
2001992013	1992-08-24	1992-08-24	anti fascists, anti fascists, antifascist, Bethnal Green, British National Party	left
2001992015	1992-10-18	1992-10-18	Cheltenham, miner, miners, pit closures	left
2001992016	1992-10-21	1992-10-25	coal miner, coal miners, coal mines, miners	left
2001992019	1992-11-15	1992-11-15	Whitburn, 636 mile, Bathgate, Glasgow to London	left
2001992020	1992-12-11	1992-12-11	soldiers, former servicemen, battalions, defence cuts	right

Table D.1: Employed Subset of the MMP Data by Clark and Regan (2019) with Additional and Recoded Variables *(continued)*

id	date	date_end	keywords	ideology
2001992021	1992-12-11	1992-12-11	fishermen, Sea Fish Conservation Bill, Sea Conservation Bill, fisheries policy, trawlermen, trawlers	right
2001993001	1993-01-17	1993-01-17	British Referendum, Maastricht, Euro sceptics, Trafalgar Square	right
2001993002	1993-03-09	1993-03-09	Peterhead harbour, Grimsby, Humberside, fishermen, fish prices, trawlermen, dock workers	left
2001993004	1993-03-22	1993-03-22	Teesport, fishermen, trawlers	right
2001993005	1993-03-27	1993-03-27	fishermen, fishing boats, fishing vessels, trawlermen	right
2001993006	1993-03-30	1993-03-30	Trades Union Congress, miners, Trotskyist Trade Union, coal	left
2001993008	1993-04-14	1993-04-14	Liverpool, fishing boats, docks	right
2001993009	1993-04-19	1993-04-19	Stornoway, council tax	left
2001993010	1993-04-25	1993-04-25	British National Party, BNP, extreme right	right
2001993011	1993-05-17	1993-05-17	Dundee, Arthur Scargill, Union of Mineworkers	left
2001993012	1993-05-23	1993-05-23	Twyford Down, M3, Dongas, chained, construction	left
2001993014	1993-05-29	1993-05-31	Kyle of Lochalsh, fishermen, fishing vessels	right
2001993016	1993-07-04	1993-07-04	Hampshire, motorway construction, M3, Twyford Down	left
2001993017	1993-07-20	1993-07-20	Wembley, policemen, Sheehy	neither
2001993018	1993-08-03	1993-08-03	Hornsey, Gardner	left
2001993019	1993-08-07	1993-08-07	Gardner, Socialist Workers	left
2001993020	1993-08-20	1993-08-20	University College Hospital, nurses, emergency cover	left

Table D.1: Employed Subset of the MMP Data by Clark and Regan (2019) with Additional and Recoded Variables *(continued)*

id	date	date_end	keywords	ideology
2001993021	1993-09-10	1993-09-10	anti racism, teenage Asian, Royal London Hospital	left
2001993022	1993-10-20	1993-10-20	pensioners, retired	left
2001993023	1993-11-05	1993-11-05	civil servants, market testing programme	left
2001994001	1994-03-05	1994-03-05	nuclear power, Anti nuclear, nuclear waste	left
2001994002	1994-04-05	1994-04-05	Walworth police station, Richard O Brien, Richard O'Brien	left
2001994003	1994-04-15	1994-04-15	Wanstead Against the M11, M11, Cambridge Park	left
2001994005	1994-07-24	1994-07-24	Criminal Justice and Public Order Bill, Criminal Justice, Justice Bill	left
2001994006	1994-10-09	1994-10-19	Criminal Justice and Public Order Bill, Criminal Justice, Justice Bill	left
2001994009	1994-11-01	1994-11-02	Means Test, means testing	left
2001994010	1994-11-03	1994-11-03	civil rights protesters, M11, Criminal Justice and Public Order Bill, Criminal Justice, Justice Bill	left
2001994011	1994-11-17	1994-11-17	Kani Yilmaz, Kurds, Kurdish	neither
2001994012	1994-11-27	1994-11-29	M11, Claremont Road, Leytonstone	left
2001994013	1994-12-05	1994-12-05	Heathrow, noise from aircraft landing, airport	left
2001995001	1995-01-09	1995-01-09	livestock, ANIMAL ACTIVISTS, Animal rights, Shoreham	left
2001995002	1995-02-14	1995-02-16	M77, Road protestors	left
2001995004	1995-02-24	1995-02-24	Brightlingsea, children, sheep, livestock, animal activists, animal rights	left

Table D.1: Employed Subset of the MMP Data by Clark and Regan (2019) with Additional and Recoded Variables *(continued)*

id	date	date_end	keywords	ideology
2001995005	1995-03-03	1995-03-03	country sports, hunting ban, British field sports, anti-hunt Bill, ban on hunting, fox hunt, fox hunting, fox hunts, ban on hunting, hunting ban, British field sports, anti-hunt Bill, blood sport	right
2001995006	1995-03-09	1995-03-09	Plymouth, animal welfare protestors, animal activists, animal rights, calves, lambs	left
2001995007	1995-05-08	1995-05-08	Stonehenge, RAMBLERS, travellers	left
2001995008	1995-05-14	1995-05-14	anti car, car free space, Reclaim The Streets	left
2001995009	1995-06-20	1995-06-20	anti roads, M11	left
2001995010	1995-07-23	1995-07-23	anti car, car free space, Reclaim The Streets	left
2001995011	1995-09-24	1995-09-24	RAMBLERS, travellers	left
2001996001	1996-01-01	1996-02-07	Newbury, A34, treehouses, tunnels, Tree Pixie Village and Granny Ash	left
2001996002	1996-02-11	1996-02-11	Newbury, Friends of the Earth, A34	left
2001996003	1996-02-16	1996-02-16	unemployed	left
2001996006	1996-05-05	1996-05-05	Wandsworth, land rights, Land is Ours	left
2001996007	1996-05-31	1996-05-31	Aberystwyth, University of Wales, Welsh Language Society	left
2001996014	1996-09-16	1996-09-16	McDonald, McDonald’s, Portillo	left
2001996016	1996-10-02	1996-10-02	Shepton Mallet, HOGG, Agriculture Minister, farmers, dairy show, National Farmers Union, Country Landowners Association	left

Table D.1: Employed Subset of the MMP Data by Clark and Regan (2019) with Additional and Recoded Variables *(continued)*

id	date	date_end	keywords	ideology
2001996017	1996-10-23	1996-10-23	farmers, National Farmers Union, BSE, HOGG, Agriculture Minister, farmers	left
2001996018	1996-10-28	1996-10-28	Northamptonshire, Heseltine, mining	left
2001996019	1996-11-03	1996-11-03	Wolverhampton, gun legislation, British Shooting Sports, gun control	right
2001996020	1996-11-19	1996-11-19	Oxford, Cambridge, higher education, UNIVERSITY, college, universities	left
2001997001	1997-01-24	1997-01-24	Northallerton, Hague, Welsh Secretary	left
2001997002	1997-02-08	1997-02-08	Chesire, ANIMAL welfare, Animal rights, fox hunt, fox hunting, fox hunts, ban on hunting, hunting ban, British field sports, anti-hunt Bill, blood sport	left
2001997003	1997-02-16	1997-02-16	Right to Work, wives, children, unemployed, women	left
2001997004	1997-02-23	1997-02-23	Sportsman s Association, Sportsman's Association, gun control, gun legislation	left
2001997005	1997-04-14	1997-04-14	Goldsmith, Referendum Party, European Union, withdrawal	right
2001997006	1997-04-18	1997-04-18	Fish, fishermen, fisheries policy, trawlermen, trawlers	right
2001997007	1997-07-10	1997-07-10	HESELTINE, ban on hunting, hunting ban, British field sports, anti-hunt Bill, ban on hunting, fox hunt, fox hunting, fox hunts	right
2001997008	1997-08-20	1997-08-20	Montserrat, Savage	neither
2001997009	1997-09-01	1997-09-01	Arms Trade, Hampshire, Farnborough, arms fair	left

Table D.1: Employed Subset of the MMP Data by Clark and Regan (2019) with Additional and Recoded Variables *(continued)*

id	date	date_end	keywords	ideology
2001997010	1997-09-28	1997-09-28	Brighton, tuition fees, BLUNKETT, university, students	left
2001997011	1997-11-26	1997-11-26	tuition fees, students, Hyde Park, university fees	left
2001997012	1997-12-01	1997-12-02	Welsh, farmers, beef, Fishguard, harbour	neither
2001997014	1997-12-26	1997-12-26	Sandringham, HUNT ENTHUSIASTS, hounds, hound, League Against Cruel Sports, blood sport, ANIMAL welfare, Animal rights, fox hunt, fox hunting, fox hunts, ban on hunting, hunting ban, British field sports, anti-hunt Bill	right
2001998001	1998-01-20	1998-01-20	farmers, Mayflower	left
2001998002	1998-01-30	1998-01-30	Catholics, Protestants, Belfast, sectarian	left
2001998005	1998-03-01	1998-03-01	Countryside Alliance, fox hunt, fox hunting, fox hunts, ban on hunting, hunting ban, British field sports, anti-hunt Bill	right
2001998006	1998-06-15	1998-06-15	Cardiff, farmers, Cunningham, Agriculture Minister, farming union	right
2001998013	1998-09-28	1998-09-28	farmers, Nick Brown, Agriculture Minister, Blackpool	right
2001998014	1998-09-28	1998-09-28	hard Left, government cuts, public services, teachers, hospital workers, trade unionists, Winter Gardens conference centre, factory closures, cutting support, minimum wage	left

Table D.1: Employed Subset of the MMP Data by Clark and Regan (2019) with Additional and Recoded Variables *(continued)*

id	date	date_end	keywords	ideology
2001998016	1998-10-02	1998-10-02	Wales, West Country, Chepstow, Winsford, farmers, lamb prices, beef imports, Tesco	neither
2001998017	1998-10-20	1998-10-20	BBC staff, National Union of Journalists, Bectu	neither
2001998020	1998-12-18	1998-12-18	Iraq, Socialist Workers Party, Al Muharjiroun, National Union of Students	left
2001999001	1999-02-25	1999-02-25	fuel duty, fuel taxes, fuel protest, fuel prices	right
2001999002	1999-03-10	1999-03-10	Falklands, anti british demonstrators	neither
2001999003	1999-03-17	1999-03-18	Portadown, Lurgan, Nelson	left
2001999004	1999-03-23	1999-03-23	truck drivers, diesel fuel, license fees	right
2001999005	1999-04-12	1999-04-12	truck drivers, diesel fuel, license fees	right
2001999006	1999-07-01	1999-07-01	Edinburgh, Northern Irish	neither
2001999008	1999-09-28	1999-09-28	Countryside Alliance, fox hunt, fox hunting, fox hunts, ban on hunting, hunting ban, British field sports, anti-hunt Bill	right
2001999009	1999-10-18	1999-10-18	unemployment, unemployed	left
2001999010	1999-11-10	1999-11-10	countryside, HUNT ENTHUSIASTS, pro hunt, hounds, hound, League Against Cruel Sports, blood sport, ANIMAL welfare, Animal rights, fox hunt, fox hunting, fox hunts, ban on hunting, hunting ban, British field sports, anti-hunt Bill	right
2001999012	1999-11-25	1999-11-25	tuition fees, university fees, student loan, student loans	left

Table D.1: Employed Subset of the MMP Data by Clark and Regan (2019) with Additional and Recoded Variables *(continued)*

id	date	date_end	keywords	ideology
2001999013	1999-11-30	1999-11-30	anti capitalist, capitalism, privatisation, anarchist, Socialist, Euston	left
2001999014	1999-12-05	1999-12-05	Protestant pupils, McGuinness	neither
2001999015	1999-12-27	1999-12-27	Essex, Gloucestershire, Wiltshire, hunters, fox hunt, fox hunting, fox hunts, ban on hunting, hunting ban, British field sports, anti-hunt Bill	right
2002000001	2000-01-11	2000-01-11	Edinburgh, farmers	right
2002000003	2000-02-14	2000-02-14	Argyll, Trident, Ploughshares, Nuclear, Disarmament	left
2002000004	2000-03-08	2000-03-08	Pankhurst, Suffrage, Suffragist	left
2002000005	2000-03-23	2000-03-23	dairy, farmers, Sainsbury, milk	right
2002000006	2000-03-29	2000-03-29	Paisley, dairy, farmers, milk	right
2002000007	2000-05-01	2000-05-01	May Day, capitalism, ANTI-CAPITALIST, GUERRILLA GARDENERS	left
2002000008	2000-05-08	2000-05-08	Midlothian, genetically modified crops, Boghall Farm, farm trials	left
2002000009	2000-05-18	2000-05-18	Tyson	left
2002000010	2000-06-07	2000-06-07	pig farmers	right
2002000011	2000-06-12	2000-06-12	Countryside Alliance, fox hunt, fox hunting, fox hunts, ban on hunting, hunting ban, British field sports, anti-hunt Bill, pro hunting	right
2002000013	2000-07-09	2000-07-09	hunting ban, hunters, fox hunt, fox hunting, foxhunting, fox hunts, ban on hunting, British field sports, anti-hunt Bill, pro hunting	right

Table D.1: Employed Subset of the MMP Data by Clark and Regan (2019) with Additional and Recoded Variables *(continued)*

id	date	date_end	keywords	ideology
2002000014	2000-08-01	2000-08-01	Argyll and Bute, Clyde, anti nuclear, Faslane, Trident, submarine base	left
2002000016	2000-08-10	2000-08-10	hunting ban, hunters, fox hunt, fox hunting, foxhunting, fox hunts, ban on hunting, British field sports, anti-hunt Bill, pro hunting	right
2002000017	2000-09-07	2000-09-15	farm workers, taxi drivers, Stanlow, Elf, refinery, Milford Haven, Texaco, Pembroke, fuel depot, Fina, Esso, Avonmouth, Bristol, oil terminal, Kingsbury, North Killingholme, Immingham, Humber Estuary, Jarrow, Fuels Terminal, fuel prices	right
2002000018	2000-11-13	2000-11-14	Edinburgh, hauliers, farmers, fuel taxes, Truck drivers, fuel protest, fuel prices, fuel duty	right
2002001001	2001-01-18	2001-01-18	hunting ban, hunters, fox hunt, fox hunting, foxhunting, fox hunts, ban on hunting, British field sports, anti-hunt Bill, pro hunting	right
2002001002	2001-01-22	2001-01-22	Balham, paedophiles	right
2002001003	2001-03-01	2001-03-01	Edinburgh, fishermen, Fish, fisheries policy, trawlermen, trawlers	right
2002001004	2001-03-12	2001-03-12	Peterhead, fishermen, Fish, fisheries policy, trawlermen, trawlers	neither
2002001005	2001-05-01	2001-05-01	May Day, capitalism, ANTI-CAPITALIST, anarchists	left
2002001006	2001-06-01	2001-06-01	Fuel protest, price of petrol, fuel taxes, fuel prices, fuel duty	right
2002001007	2001-06-06	2001-06-06	Harehills, Asian youths, Asian man	left

Table D.1: Employed Subset of the MMP Data by Clark and Regan (2019) with Additional and Recoded Variables *(continued)*

id	date	date_end	keywords	ideology
2002001008	2001-07-20	2001-07-20	Brixton, fatal shooting	left
2002001009	2001-08-05	2001-08-06	asylum seekers	left
2002001012	2001-08-07	2001-08-07	pool, Govanhill	left
2002001013	2001-08-21	2001-08-21	Edinburgh, red light district	right
2002001014	2001-09-11	2001-09-11	arms fair, arms trade, defence systems	left
2002001015	2001-11-18	2001-11-18	anti war, anti-war, bradshaw, PEACE DEMONSTRATION, PEACE PROTEST, ANTI-TERROR, ANTI TERROR, STOP THE WAR, WAR ON TERROR, Iraq	left
2002001017	2001-12-10	2001-12-10	hunting ban, hunters, fox hunt, fox hunting, foxhunting, fox hunts, ban on hunting, British field sports, anti-hunt Bill, pro hunting, Edinburgh, Countryside Alliance, Rural rebels, hunting with dogs	right
2002002001	2002-01-18	2002-01-18	Belfast, Londonderry, Omagh, Enniskillen, Newry, Cookstown, Strabane, Northern Ireland, sectarian violence, paramilitary groups	left
2002002002	2002-02-06	2002-02-07	hunting ban, hunters, fox hunt, fox hunting, foxhunting, fox hunts, ban on hunting, British field sports, anti-hunt Bill, pro hunting, Countryside Alliance, Rural rebels, riders	right
2002002004	2002-02-11	2002-02-11	Argyll, anti nuclear, Faslane, Trident, submarine base, nuclear deterrent, nuclear weapons, vanguard, nuclear protest	left

Table D.1: Employed Subset of the MMP Data by Clark and Regan (2019) with Additional and Recoded Variables *(continued)*

id	date	date_end	keywords	ideology
2002002005	2002-02-13	2002-02-13	Edinburgh, hunting ban, hunters, fox hunt, fox hunting, foxhunting, fox hunts, ban on hunting, British field sports, anti-hunt Bill, pro hunting, Countryside Alliance, Rural rebels, hunting with dogs	left
2002002006	2002-03-13	2002-03-13	policemen, rank and file, officers, reforms to the service, Police Force, POLICE protest	neither
2002002009	2002-05-01	2002-05-01	May Day, capitalism, ANTI-CAPITALIST, anarchists, mayday	left
2002002010	2002-05-19	2002-05-19	Worcestershire, asylum seekers	right
2002002011	2002-05-22	2002-05-22	Country Sports, Hounds, foxhunting, hunting ban, hunters, fox hunt, fox hunting, foxhunting, fox hunts, ban on hunting, British field sports, anti-hunt Bill, pro hunting, Countryside Alliance, Rural rebels, hunting with dogs	right
2002002012	2002-06-04	2002-06-04	anti monarch, anti-monarch, anti monarchy, anti-monarchy, anti-royalists	left
2002002016	2002-07-12	2002-07-12	Doncaster	right
2002002018	2002-09-10	2002-09-10	Gibraltar, National Day	neither
2002002019	2002-09-16	2002-09-23	hunting ban, hunters, fox hunt, fox hunting, foxhunting, fox hunts, ban on hunting, British field sports, anti-hunt Bill, pro hunting, Countryside Alliance, Rural rebels, rural protesters	right

Table D.1: Employed Subset of the MMP Data by Clark and Regan (2019) with Additional and Recoded Variables *(continued)*

id	date	date_end	keywords	ideology
2002002021	2002-10-05	2002-10-06	Belfast, Sinn Fein, Irish Republican Army, IRA	neither
2002002023	2002-10-28	2002-10-28	Firefighters, fireman, firewoman, firefighting	neither
2002002025	2002-11-01	2002-11-01	anti war, anti-war, antiwar, PEACE DEMONSTRATION, PEACE PROTEST, ANTI-TERROR, ANTI TERROR, STOP THE WAR, WAR ON TERROR, Iraq	left
2002002027	2002-11-15	2002-11-15	cruise missiles, Greenham Common, Berkshire, anti-cruise	left
2002002028	2002-11-24	2002-11-24	Westminster Bridge, student loan, student loans, Socialist Worker, tuition fees, university fees	left
2002002030	2002-12-08	2002-12-08	Firefighters, fireman, firewoman, firefighting	neither
2002002031	2002-12-11	2002-12-11	fishing boats, fishermen, Fish, fisheries policy, trawlermen, trawlers, English Channel	right
2002002032	2002-12-17	2002-12-17	Country Sports, Hounds, foxhunting, hunting ban, hunters, fox hunt, fox hunting, foxhunting, fox hunts, ban on hunting, British field sports, anti-hunt Bill, pro hunting, Countryside Alliance, Rural rebels, hunting with dogs	right

Table D.1: Employed Subset of the MMP Data by Clark and Regan (2019) with Additional and Recoded Variables *(continued)*

id	date	date_end	keywords	ideology
2002002033	2002-12-26	2002-12-26	Boxing Day hunt, Country Sports, Hounds, foxhunting, hunting ban, hunters, fox hunt, fox hunting, foxhunting, fox hunts, ban on hunting, British field sports, anti-hunt Bill, pro hunting, Countryside Alliance, Rural rebels, hunting with dogs	right
2002003001	2003-01-18	2003-01-19	Northwood, military headquarters	left
2002003002	2003-01-20	2003-01-20	Lincolnshire, Sittingbourne, Kent, asylum seekers	right
2002003003	2003-01-23	2003-01-23	anti war, anti-war, antiwar, PEACE DEMONSTRATION, PEACE PROTEST	left
2002003004	2003-02-15	2003-02-15	anti war, anti-war, antiwar, PEACE DEMONSTRATION, PEACE PROTEST, ANTI-TERROR, ANTI TERROR, STOP THE WAR, WAR ON TERROR, Iraq	left
2002003006	2003-03-05	2003-03-05	anti war, anti-war, antiwar, PEACE DEMONSTRATION, PEACE PROTEST	left
2002003007	2003-03-16	2003-03-23	anti war, anti-war, antiwar, PEACE DEMONSTRATION, PEACE PROTEST, ANTI-TERROR, ANTI TERROR, STOP THE WAR, WAR ON TERROR, Iraq	left

Table D.1: Employed Subset of the MMP Data by Clark and Regan (2019) with Additional and Recoded Variables *(continued)*

id	date	date_end	keywords	ideology
2002003010	2003-03-22	2003-03-22	anti war, anti-war, antiwar, PEACE DEMONSTRATION, PEACE PROTEST, ANTI-TERROR, ANTI TERROR, STOP THE WAR, WAR ON TERROR, Iraq	left
2002003013	2003-04-13	2003-04-13	anti war, anti-war, antiwar, PEACE DEMONSTRATION, PEACE PROTEST, ANTI-TERROR, ANTI TERROR, STOP THE WAR, WAR ON TERROR, Iraq	left
2002003014	2003-04-22	2003-04-22	Argyll and Bute, CND, Faslane, Trident, Bruce Kent, nuclear disarmament, nuclear deterrent, nuclear weapons, anti nuclear, anti-nuclear	left
2002003015	2003-05-01	2003-05-01	May Day, capitalism, ANTI-CAPITALIST, anarchists, mayday	left
2002003016	2003-06-29	2003-06-29	Country Sports, Hounds, foxhunting, hunting ban, hunters, fox hunt, fox hunting, foxhunting, fox hunts, ban on hunting, British field sports, anti-hunt Bill, pro hunting, Countryside Alliance, Rural rebels, hunting with dogs	right
2002003017	2003-09-27	2003-09-27	anti war, anti-war, antiwar, PEACE DEMONSTRATION, PEACE PROTEST, ANTI-TERROR, ANTI TERROR, STOP THE WAR, WAR ON TERROR, Iraq	left

Table D.1: Employed Subset of the MMP Data by Clark and Regan (2019) with Additional and Recoded Variables *(continued)*

id	date	date_end	keywords	ideology
2002003018	2003-10-06	2003-10-06	pensioners, pensions credit, retired	left
2002003019	2003-10-26	2003-10-26	student loan, student loans, tuition fees, university fees	left
2002003020	2003-11-19	2003-11-20	anti war, anti-war, antiwar, PEACE DEMONSTRATION, PEACE PROTEST, ANTI-TERROR, ANTI TERROR, STOP THE WAR, WAR ON TERROR, Iraq, Bush	left
2002004001	2004-02-05	2004-02-05	relatives were murdered, loyalist paramilitaries, M15	neither
2002004002	2004-02-17	2004-02-17	Kensington, Chelsea, congestion charge, congestion fee, congestion fees, fees for congestion	neither
2002004003	2004-03-01	2004-03-01	pensioners, pension bill, retired, pension reform	left
2002004004	2004-03-18	2004-03-18	Kurdish, asylum seekers, asylum seeker, hunger strike, refugees, refugee	left
2002004005	2004-04-02	2004-04-02	animal activists, hedgehogs, Duchess of Hamilton, ANIMAL welfare, Animal rights	left
2002004006	2004-04-09	2004-04-09	National Front, Muslims, Finsbury Park mosque, imam, Abu Hamza	right
2002004008	2004-06-15	2004-06-15	lorry drivers, fuel prices, environmentalists, climate change, truck drivers, environmental activists, fossil fuels, petrol prices	right
2002004009	2004-07-01	2004-07-01	Eros, Piccadilly Circus, Mick Jagger, Keith Richards	left

Table D.1: Employed Subset of the MMP Data by Clark and Regan (2019) with Additional and Recoded Variables *(continued)*

id	date	date_end	keywords	ideology
2002004012	2004-08-23	2004-08-23	Argyll and Bute, CND, Faslane, Trident, Bruce Kent, nuclear disarmament, nuclear deterrent, nuclear weapons, anti nuclear, anti-nuclear	left
2002004013	2004-09-15	2004-09-15	Country Sports, Hounds, foxhunting, hunting ban, hunters, fox hunt, fox hunting, foxhunting, fox hunts, ban on hunting, British field sports, anti-hunt Bill, pro hunting, Countryside Alliance, Rural rebels, hunting with dogs	right
2002004014	2004-09-28	2004-09-28	Brighton, Country Sports, Hounds, foxhunting, hunting ban, hunters, fox hunt, fox hunting, foxhunting, fox hunts, ban on hunting, British field sports, anti-hunt Bill, pro hunting, Countryside Alliance, Rural rebels, hunting with dogs, anti war, anti-war, antiwar, PEACE DEMONSTRATION, PEACE PROTEST, ANTI-TERROR, ANTI TERROR, STOP THE WAR, WAR ON TERROR, Iraq, Bush	right
2002004015	2004-10-17	2004-10-17	anti war, anti-war, antiwar, PEACE DEMONSTRATION, PEACE PROTEST, ANTI-TERROR, ANTI TERROR, STOP THE WAR, WAR ON TERROR, Iraq, Bush	left

Table D.1: Employed Subset of the MMP Data by Clark and Regan (2019) with Additional and Recoded Variables *(continued)*

id	date	date_end	keywords	ideology
2002004017	2004-11-18	2004-11-18	Berkshire, Country Sports, Hounds, foxhunting, hunting ban, hunters, fox hunt, fox hunting, foxhunting, fox hunts, ban on hunting, British field sports, anti-hunt Bill, pro hunting, Countryside Alliance, Rural rebels, hunting with dogs	right
2002005001	2005-02-19	2005-02-19	Country Sports, Hounds, foxhunting, hunting ban, hunters, fox hunt, fox hunting, foxhunting, fox hunts, ban on hunting, British field sports, anti-hunt Bill, pro hunting, Countryside Alliance, Rural rebels, hunting with dogs	right
2002005002	2005-03-07	2005-03-07	anti terrorism, terrorism Act, British Institute of Human Rights, liberty	left
2002005003	2005-04-25	2005-04-26	haulers, oil refineries, fuel prices, Fuel protest, price of petrol, fuel taxes, fuel duty, cost of fuel	right
2002005005	2005-07-06	2005-07-07	Auchterarder, Stirling, Anti war protest, anti poverty, G8, anti-war, antiwar, PEACE DEMONSTRATION, PEACE PROTEST, ANTI-TERROR, ANTI TERROR, STOP THE WAR, WAR ON TERROR	left
2002005006	2005-07-08	2005-07-08	anti capitalist, Dissent, M74, Glasgow	left
2002005008	2005-08-07	2005-08-07	Serious and Organised Crime and Police Act	left

Table D.1: Employed Subset of the MMP Data by Clark and Regan (2019) with Additional and Recoded Variables *(continued)*

id	date	date_end	keywords	ideology
2002005009	2005-08-16	2005-08-16	farmers, Brazilian beef, low cost beef, beef industry	neither
2002005012	2005-11-12	2005-11-12	pensioners, NPC, pensions, healthcare, retired	left
2002005013	2005-11-21	2005-11-21	church leaders, Tommy Sheridan, Paddy Hill, immigration offices, dawn raids, asylum seekers	left
2002005014	2005-12-18	2005-12-18	Prestwick, human rights, airports, CIA torture, Stop the War	left
2002006001	2006-01-16	2006-01-16	BNP, Nick Griffin, racial hatred, anti fascist, British National Party, antifascist	right
2002006003	2006-02-01	2006-02-01	anti war, anti-war, antiwar, PEACE DEMONSTRATION, PEACE PROTEST, ANTI-TERROR, ANTI TERROR, STOP THE WAR, WAR ON TERROR, Iraq	left
2002006004	2006-03-18	2006-03-18	anti war, anti-war, antiwar, PEACE DEMONSTRATION, PEACE PROTEST, ANTI-TERROR, ANTI TERROR, STOP THE WAR, WAR ON TERROR, Iraq	left
2002006005	2006-05-21	2006-05-21	die in, ANIMAL welfare, Animal rights, animal-rights, animal-welfare	left
2002006006	2006-06-09	2006-06-09	MUSLIMS, police station, Forest Gate	neither
2002006007	2006-09-07	2006-09-07	anti war, anti-war, antiwar, PEACE DEMONSTRATION, PEACE PROTEST, ANTI-TERROR, ANTI TERROR, STOP THE WAR, WAR ON TERROR, Iraq	left

Table D.1: Employed Subset of the MMP Data by Clark and Regan (2019) with Additional and Recoded Variables *(continued)*

id	date	date_end	keywords	ideology
2002006008	2006-09-14	2006-09-14	CND, Faslane, Trident, nuclear disarmament, nuclear deterrent, nuclear weapons, anti nuclear, anti-nuclear	left
2002006010	2006-09-19	2006-09-19	CND, Faslane, Trident, nuclear disarmament, nuclear deterrent, nuclear weapons, anti nuclear, anti-nuclear	left
2002006011	2006-10-25	2006-10-25	Pensioners, retired, pension	left
2002006012	2006-10-29	2006-10-29	students, Trafalgar Square, top up fees, student loan, student loans, tuition fees, university fees	left
2002006013	2006-12-10	2006-12-10	Glasgow, positive action, dawn raids, asylum seekers, immigration centre, Govan	left
2002006014	2006-12-11	2006-12-11	Somerset, pedophile	right
2002007001	2007-01-07	2007-01-07	weapons of mass destruction, naval base yesterday, Argyll and Bute, CND, Faslane, Trident, Bruce Kent, nuclear disarmament, nuclear deterrent, nuclear weapons, anti nuclear, anti-nuclear	left
2002007002	2007-01-09	2007-01-09	Christians, hymns, prayed, sexual orientation	right
2002007003	2007-05-07	2007-05-07	Argyll and Bute, CND, Faslane, Trident, Bruce Kent, nuclear disarmament, nuclear deterrent, nuclear weapons, anti nuclear, anti-nuclear	left

Table D.1: Employed Subset of the MMP Data by Clark and Regan (2019) with Additional and Recoded Variables *(continued)*

id	date	date_end	keywords	ideology
2002007004	2007-07-01	2007-07-01	Bolton, smokers, cigarettes, pipes, lighters, The Swan, Nick Hogan, ban on smoking	right
2002007005	2007-08-17	2007-08-17	camping, anti-war, against the wars, anti war, anti-war, antiwar, PEACE DEMONSTRATION, PEACE PROTEST, ANTI-TERROR, ANTI TERROR, STOP THE WAR, WAR ON TERROR, Iraq	left
2002007007	2007-10-01	2007-10-01	clowns, Argyll and Bute, CND, Faslane, Trident, Bruce Kent, nuclear disarmament, nuclear deterrent, nuclear weapons, anti nuclear, anti-nuclear	left
2002007008	2007-10-08	2007-10-08	anti war, anti-war, antiwar, PEACE DEMONSTRATION, PEACE PROTEST, ANTI-TERROR, ANTI TERROR, STOP THE WAR, WAR ON TERROR, Iraq	left
2002008001	2008-01-23	2008-01-23	police, policemen, rank and file, officers, reforms to the service, Police Force, POLICE protest	neither
2002008002	2008-04-20	2008-04-20	Ethnic Catering Alliance, immigration, ethnic catering	left
2002008003	2008-04-29	2008-04-29	haulers, oil refineries, fuel prices, Fuel protest, price of petrol, fuel taxes, fuel duty, fuel protesters, petrol prices, cost of fuel	right
2002008004	2008-05-01	2008-05-01	Edinburgh, students, Die Meistersinger	left

Table D.1: Employed Subset of the MMP Data by Clark and Regan (2019) with Additional and Recoded Variables *(continued)*

id	date	date_end	keywords	ideology
2002008005	2008-05-27	2008-05-27	haulers, oil refineries, fuel prices, Fuel protest, price of petrol, fuel taxes, fuel duty, fuel protesters, lorry drivers, petrol prices, cost of fuel	right
2002008006	2008-06-10	2008-06-16	Edinburgh, Glasgow, haulers, oil refineries, fuel prices, Fuel protest, price of petrol, fuel taxes, fuel duty, fuel protesters, lorry drivers, petrol prices, trucks, cost of fuel	right
2002008008	2008-09-24	2008-09-24	Edinburgh, Aberdeen, Dumfries, Inverness, Unison, GMB, Unite Unions, local government workers, Princes Street Gardens	left
2002008009	2008-11-02	2008-11-02	Belfast	neither
2002008010	2008-12-04	2008-12-04	Edinburgh, Aisha al Megrahi, Lockerbie bomber	neither
2002009001	2009-01-12	2009-01-12	Climate rush, environmental, Terminal 1, Heathrow, picnic, climate change	left
2002009002	2009-01-30	2009-01-30	WILDCAT, foreign contractors, Fiddlers Ferry, Warrington, British jobs, British workers	right
2002009003	2009-02-03	2009-02-03	Construction workers, Longannet, Cockenzie, power stations, Lindsay, oil refinery, Lincolnshire, foreign workers	neither
2002009004	2009-02-05	2009-02-05	taxi drivers, black cabs, black cab, minicabs	neither
2002009005	2009-02-16	2009-02-16	photographers	left
2002009006	2009-03-07	2009-03-07	Glasgow, primary schools, nurseries, Newark Drive, Nithsdale Road	left
2002009007	2009-03-08	2009-03-08	Tibetans, Tibet	neither

Table D.1: Employed Subset of the MMP Data by Clark and Regan (2019) with Additional and Recoded Variables *(continued)*

id	date	date_end	keywords	ideology
2002009008	2009-03-10	2009-03-10	Muslim anti war, Luton, The Royal Anglian Regiment, The Poachers	left
2002009016	2009-04-01	2009-04-01	G20, anti capitalist, anticapitalist, financial district	left
2002009017	2009-05-05	2009-05-05	savers, Bank of England	neither
2002009018	2009-06-09	2009-06-09	UAF, anti racism, anti-racism, antiracism, Griffin, racial hatred, BNP, B.N.P., British National Party	left
2002009019	2009-08-28	2009-09-03	Blackheath, Climate Camp, climate change, coal fired, power station, Great Climate Swoop	left
2002009020	2009-10-18	2009-10-18	Nottinghamshire, climate change, coal fired, power station	left
2002009021	2009-10-22	2009-10-22	anti racism, anti-racism, antiracism, Griffin, BNP, B.N.P., British National Party	left
2002010001	2010-01-29	2010-01-29	anti war, anti-war, antiwar, STOP THE WAR, WAR ON TERROR, Iraq	left
2002010002	2010-04-22	2010-04-22	anti war, anti-war, antiwar, STOP THE WAR, WAR ON TERROR, Iraq	left
2002010003	2010-05-01	2010-07-20	Democracy Village, Afghanistan	left
2002010004	2010-06-21	2010-06-21	Academics, students, funding cuts, tuition fees, student loan, student loans, tuition fees, university fees	left
2002010005	2010-08-24	2010-08-24	Camp for Climate Action, fossil fuel, climate change, environmental, environment	left
2002010006	2010-09-10	2010-09-10	artists, Tate Modern, arts spending	left
2002010008	2010-10-23	2010-10-23	Scottish Trades Union, public spending cuts, Government cuts	left

Table D.1: Employed Subset of the MMP Data by Clark and Regan (2019) with Additional and Recoded Variables *(continued)*

id	date	date_end	keywords	ideology
2002010009	2010-10-27	2010-10-27	Vodafone	left
2002010011	2010-11-10	2010-11-10	education, tuition, university, funding cuts, student loan	left
2002010012	2010-11-24	2010-12-08	Leeds	left
2002010013	2010-11-24	2010-11-24	student protesters, students, tuition fees, funding cuts, student loan, student loans, university fees	left
2002010014	2010-11-30	2010-11-30	student protesters, Trafalgar Square, tuition fees, funding cuts, student loan, student loans, university fees	left
2002010015	2010-12-04	2010-12-04	UK uncut, tax avoidance, TAX DODGERS, Tax protest, corporate tax	left
2002010016	2010-12-09	2010-12-09	student protesters, tuition fees, student loan, student loans, university fees	left
2002011001	2011-01-29	2011-01-29	Manchester, student protesters, Trafalgar Square, tuition fees, funding cuts, student loan, student loans, university fees	left
2002011002	2011-02-17	2011-02-17	Glasgow, university students, faculty, University of Glasgow, funding cuts	left
2002011003	2011-03-21	2011-03-21	anti war, anti-war, antiwar, STOP THE WAR, WAR ON TERROR, Iraq	left
2002011004	2011-03-22	2011-03-22	student protesters, higher education, Holyrood, RIP uni, tuition fees, NUS Scotland, Reclaim Your Voice, student loan, student loans, university fees	left

Table D.1: Employed Subset of the MMP Data by Clark and Regan (2019) with Additional and Recoded Variables *(continued)*

id	date	date_end	keywords	ideology
2002011005	2011-03-26	2011-03-26	Trades Union, austerity measures, TUC, cuts in public spending, austerity, anti-cuts	left
2002011006	2011-04-22	2011-04-22	Bristol, squatters, squat housing, Whose streets	left
2002011007	2011-04-29	2011-04-29	Bristol, masked protesters, Tesco, Cheltenham Road, Whose streets	left
2002011008	2011-06-22	2011-06-22	Strathclyde, students, university staff, Hetherington Research Centre, cuts	left
2002011009	2011-06-30	2011-06-30	teachers, public sector workers, austerity measures, pension plans	left
2002011011	2011-10-15	2011-11-05	Occupy, 99 percent, City financial district, Stock Exchange, anti capitalist, anti-capitalist, anticapitalist, Bath, Belfast, Birmingham, Bournemouth, Bradford, Brighton, Bristol, Cardiff, Edinburgh, Exeter, Glasgow, Lampeter, Lancaster, Leeds, Liverpool, London, Manchester, Norwich, Plymouth, Sheffield, Thanet, University of Brighton, University of Warwick	left
2002011013	2011-10-19	2011-10-19	Basildon, travelers, Dale Farm	left
2002011015	2011-11-09	2011-11-09	student protesters, tuition fees, student loan, student loans, university fees	left
2002011018	2011-11-30	2011-11-30	Scotland for Marriage, anti gay, anti-gay, antigay, marriage, Same sex	right
2002011019	2011-11-30	2011-11-30	workers, pensions, public sector strike, Graeme Smith, Scottish Trades Union	left

Table D.1: Employed Subset of the MMP Data by Clark and Regan (2019) with Additional and Recoded Variables *(continued)*

id	date	date_end	keywords	ideology
2002012001	2012-03-29	2012-03-29	Leyton, Olympic Park, Occupy, 99 percent, City financial district, Stock Exchange, anti capitalist, anti-capitalist, anticapitalist	left
2002012002	2012-04-28	2012-04-28	cyclists, bike friendly, cycling, bicycle	left
2002012004	2012-05-10	2012-05-10	police officers, civil servants, budget cuts, pension plans	neither
2002012007	2012-07-17	2012-07-17	Taxi Drivers, drivers union, traffic lanes	neither
2002012008	2012-07-27	2012-07-27	cyclists, bike friendly, cycling, bicycle	left
2002012009	2012-08-24	2012-08-24	Anti LGBT, Anti-LGBT, gay marriage,Renfrew	right
2002012012	2012-10-12	2012-10-12	Veterans, Royal Regiment of Fusiliers, military job cuts, infantry battalions	neither
2002012013	2012-10-18	2012-10-18	Anti Choice, Anti-Choice, AntiChoice, abortion, family planning, Marie Stopes, abortions	right
2002012014	2012-11-13	2012-11-13	Lawyers, legal aid system, Edinburgh Bar Association,solicitors,legal aid	left
2002012015	2012-12-03	2013-01-08	Belfast City Hall, Union flag	right
2002013005	2013-02-05	2013-02-05	gay marriage, same-sex marriage	left
2002013006	2013-03-26	2013-03-26	University of Sussex, Sussex University	left
2002013007	2013-04-16	2013-04-16	Faslane, Argyll and Bute, CND, Trident, nuclear disarmament, nuclear deterrent, nuclear weapons, anti nuclear, anti-nuclear	left
2002013008	2013-05-25	2013-05-25	Balcombe, Cuadrilla Resources, hydraulic fracturing, fracking, shale gas	left

Table D.1: Employed Subset of the MMP Data by Clark and Regan (2019) with Additional and Recoded Variables *(continued)*

id	date	date_end	keywords	ideology
2002013013	2013-07-16	2013-07-16	cyclists, bike friendly, cycling, bicycle, cycle	left
2002013014	2013-07-25	2013-08-21	Balcombe, environmentalists, environmentalist, hydraulic fracturing, fracking, shale gas, oil exploration	left
2002013015	2013-08-09	2013-08-09	Internment	neither
2002013016	2013-08-14	2013-08-14	train fares, rail fares, railway	left
2002013019	2013-10-30	2013-10-30	Stop Climate, Climate Chaos, cycling, environmentalists, environmentalist, cyclists, bike friendly, cycling, bicycle, cycle	left
2002013020	2013-11-29	2013-11-29	cyclists, bike friendly, cycling, bicycle, cycle, Transport for London, Southwark	left
2002014001	2014-01-08	2014-01-08	Mark Duggan	left
2002014002	2014-03-08	2014-03-08	barristers, solicitors, legal aid	left
2002014003	2014-06-11	2014-06-11	black cab, Taxi Drivers, cabdrivers, Uber	neither
2002014004	2014-06-29	2014-06-29	Pro independence, Pro-independence, BBC Scotland	neither
2002014005	2014-07-01	2014-07-01	Lady Gaga, Smithfield Market	neither
2002014006	2014-08-01	2014-08-01	black cab, Taxi Drivers, cabdrivers, Uber	neither
2002014007	2014-09-03	2014-09-03	Port Seton, Coastal Regeneration, CRA, Historic Scotland, Cockenzie, power station, energy park, wind farm	left
2002014008	2014-09-12	2014-09-12	anti UKIP, anti-UKIP, Farage	left
2002014009	2014-09-14	2014-09-14	Yes campaign, referendum, BBC headquarters, Nick Robinson, Yes camp, Pacific Quay	neither

Table D.1: Employed Subset of the MMP Data by Clark and Regan (2019) with Additional and Recoded Variables *(continued)*

id	date	date_end	keywords	ideology
2002014010	2014-09-21	2014-09-21	carbon emissions, climate change	left
2002014011	2014-10-18	2014-10-18	Trades Union, Pay Rise	left
2002014012	2014-11-05	2014-11-05	Anti capitalists, Anti-capitalists, Anticapitalists, Anti capitalist, Anti-capitalist, Anticapitalist, Million Mask March, Anonymous, Guy Fawkes masks	left
2002014013	2014-11-19	2014-11-19	student protesters, tuition fees, student loan, student loans, university fees	left
2002015001	2015-03-07	2015-03-07	West Drayton, asylum seekers, Harmondsworth, immigration removal	left
2002015003	2015-04-01	2015-04-01	Faslane, CND, Trident, nuclear disarmament, nuclear deterrent, nuclear weapons, anti nuclear, anti-nuclear	left
2002015004	2015-04-13	2015-04-13	Faslane, Argyll and Bute, CND, Trident, nuclear disarmament, nuclear deterrent, nuclear weapons, anti nuclear, anti-nuclear	left
2002015005	2015-04-20	2015-04-20	Cannabis, marijuana	left
2002015006	2015-04-25	2015-04-25	Brixton, gentrification, anti-gentrification	left
2002015007	2015-05-09	2015-05-09	London Black Revolutionaries, anti austerity, anti-austerity	left
2002015008	2015-05-26	2015-05-26	minicabs, black cab, Taxi Drivers, cabdrivers, Uber	neither
2002015009	2015-06-15	2015-06-15	cyclists, Dare to bare, stripped, Naked Bike Ride, bike friendly, cycling, bicycle, cycle	left

Table D.1: Employed Subset of the MMP Data by Clark and Regan (2019) with Additional and Recoded Variables *(continued)*

id	date	date_end	keywords	ideology
2002015010	2015-06-20	2015-06-20	anti austerity, anti-austerity, austerity, spending cuts, public spending	left
2002015011	2015-08-16	2015-08-16	environmentalists, environmentalist, BP, Edinburgh International Festival, Edinburgh Festival	left
2002015013	2015-10-04	2015-10-06	anti austerity, anti-austerity, austerity, spending cuts, public spending	left
2002015017	2015-11-04	2015-11-04	anarchist, black bloc, student demonstration, tuition fees,maintenance grants	left
2002015018	2015-11-05	2015-11-05	Anti capitalists, Anti-capitalists, Anticapitalists, Anti capitalist, Anti-capitalist, Anticapitalist, Million Mask March, Anonymous, Guy Fawkes masks	left
2002015019	2015-12-01	2015-12-01	Stop the War, Islamic State	left
2002016001	2016-01-02	2016-01-04	Action for Rail, train fares, rail fares, railway	left
2002016003	2016-01-27	2016-01-27	fishermen, fishing	right
2002016004	2016-01-30	2016-01-30	Dewsbury, Britain First, farright, far right, antifascist, anti fascist, pro immigrant, pro-immigrant, pro immigration, pro-immigration, anti immigration, anti-immigration, anti immigrant, anti-immigrant	right
2002016008	2016-02-24	2016-02-24	union members	left
2002016009	2016-03-10	2016-03-10	farmers, farmer	left
2002016010	2016-04-07	2016-04-07	Carnegie Library	left

Table D.1: Employed Subset of the MMP Data by Clark and Regan (2019) with Additional and Recoded Variables *(continued)*

id	date	date_end	keywords	ideology
2002016011	2016-04-16	2016-04-16	anti austerity, anti-austerity, austerity, spending cuts, public spending, teachers, nurses	left
2002016012	2016-04-27	2016-04-27	hydraulic fracturing, fracking, shale gas, oil exploration, Emma Thompson, Greenpeace, Cuadrilla	left
2002016013	2016-04-27	2016-04-27	junior doctors, Jeremy Hunt, NHS walkout	left
2002016014	2016-04-30	2016-04-30	Trident, nuclear disarmament, nuclear deterrent, nuclear weapons, anti nuclear, anti-nuclear, Nicola Sturgeon	left
2002016015	2016-06-15	2016-06-15	Thames, Farmers, Farmer, UKIP, Nigel Farage, Brexit, anti Brexit, anti-Brexit, European Union	right
2002016018	2016-06-26	2016-06-26	Birmingham, far right, far-right, farright, mosque	right
2002016019	2016-06-28	2016-06-28	EU, E.U., european union, Brexit	left
2002016020	2016-07-02	2016-07-02	EU, E.U., european union, Brexit	left
2002016022	2016-07-11	2016-07-11	black lives matter, police harassment, racial injustice, police brutality, hate crimes, Mark Duggan, black man	left
2002016023	2016-07-11	2016-07-11	Victoria station, commuters, rail network, Southern Rail	left
2002016024	2016-07-18	2016-07-18	Trident, nuclear disarmament, nuclear deterrent, nuclear weapons, anti nuclear, anti-nuclear	left
2002016025	2016-08-05	2016-08-05	black lives matter, police harassment, racial injustice, police brutality, hate crimes, Mark Duggan, black man	left
2002016026	2016-10-15	2016-10-15	Citizens UK, Shakeel Begg	left

Table D.1: Employed Subset of the MMP Data by Clark and Regan (2019) with Additional and Recoded Variables *(continued)*

id	date	date_end	keywords	ideology
2002016027	2016-11-05	2016-11-05	Million Mask March, anti government, Anti capitalists, Anti-capitalists, Anticapitalists, Anti capitalist, Anti-capitalist, Anticapitalist, Million Mask March, Anonymous, Guy Fawkes masks	left
2002016030	2016-11-15	2016-11-15	Devon, hospital closure, hospital closures	left
2002016032	2016-11-15	2016-11-15	prison officers	neither
2002016035	2016-12-20	2017-01-02	Kirby Misperton, hydraulic fracturing, fracking, shale gas	left
2002017001	2017-01-01	2017-01-01	Lake District, Duddon estuary, Askamin Furness, electricity pylons	left
2002017003	2017-01-21	2017-01-21	feminist, gender equality, women's rights, Women's March, Women's marches	left
2002017004	2017-01-30	2017-01-30	Trump, Trump's, Muslim ban	left
2002017006	2017-02-04	2017-02-04	Trump, state visit	left
2002017011	2017-03-04	2017-03-04	nurses, NHS	left
2002017012	2017-03-25	2017-03-25	Remainers, against Brexit, Pro-EU, Anti-Brexit, Pro EU, Anti Brexit, REMAIN MARCH	left
2002017013	2017-05-07	2017-05-07	stabbings, knife crime	neither
2002017014	2017-06-16	2017-06-16	Grenfell	left
2002017015	2017-06-21	2017-06-21	Shepherds Bush, Day of Rage	left
2002017016	2017-06-25	2017-06-25	Edir Frederico Da Costa, Forest Gate	left
2002017017	2017-07-28	2017-07-28	black man, restrained by police, Rashan Charles	left

Table D.1: Employed Subset of the MMP Data by Clark and Regan (2019) with Additional and Recoded Variables *(continued)*

id	date	date_end	keywords	ideology
2002017018	2017-10-01	2017-10-01	against Brexit, anti Brexit, anti-Brexit, anti austerity, anti-austerity, party conference	left
2002017019	2017-10-28	2017-10-28	Wick, Thurso, Portree, Caithness, hospital provision, hospital services, Isle of Skye	left
2002017020	2017-11-21	2017-11-21	Jordanhill Community, Jordanhill college	left

Note: Some events were merged, meaning they received the start date of the first and the end data of the last event included in the new entry.

E Additional Models

	Delegitimising Frames	Legitimising Frames
(Intercept)	0.590^{***}	-0.501^{***}
	(0.147)	(0.146)
Event-Level Factors		
goal: anti-war	-0.053	-0.156
	(0.135)	(0.142)
goal: labour protests	0.068	0.027
	(0.072)	(0.077)
goal: police	0.313	-0.250
	(0.190)	(0.204)
goal: social-issue	-0.245^{**}	0.829^{***}
	(0.076)	(0.077)
violent protest	0.325^{***}	-0.102
	(0.067)	(0.069)
state response: arrests	-0.020	-0.240^{**}
	(0.080)	(0.086)
state response: beatings	0.008	-0.220
	(0.142)	(0.151)
state response: crowd dispersal	0.019	-0.196^{*}
	(0.078)	(0.083)
state response: ignore	-0.105	0.229^{**}
	(0.082)	(0.086)
Outlet-Level Factors		
right-wing	0.009	-0.086
	(0.106)	(0.103)
tabloid newspaper	-0.104	-0.384^{***}
	(0.110)	(0.108)
ideological divide: conflict	0.081	-0.090
	(0.073)	(0.072)
ideological divide: ambiguous	0.273^{*}	0.611^{***}
	(0.131)	(0.130)
Time Bound Factors		
days since start	-0.720^{***}	-0.231
	(0.170)	(0.185)
Year of protest (since 1992)	-0.155	0.106
	(0.181)	(0.175)
AIC	13067.579	12225.669
BIC	13190.438	12348.528
Log Likelihood	-6516.789	-6095.834
Num. obs.	10168	10168
Num. groups: np_year	180	180
Var: np_year (Intercept)	0.292	0.261

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{+}p < 0.1$

Table E.1: Regression Results for Aggregated Frames (with all State Responses)

	Delegitimising Frames	Legitimising Frames
(Intercept)	0.538^{***}	-0.395^{**}
	(0.132)	(0.137)
Event-Level Factors		
goal: anti-war	0.067	−0.086
	(0.181)	(0.192)
goal: labour protests	0.057	0.266^{*}
	(0.118)	(0.119)
goal: police	0.359	−0.253
	(0.285)	(0.285)
goal: social-issue	-0.189^{*}	0.668^{***}
	(0.096)	(0.097)
violent protest	0.389^{***}	-0.458^{***}
	(0.084)	(0.085)
repression of peaceful p.	0.010	-0.498^{***}
	(0.104)	(0.109)
Outlet-Level Factors		
right-wing	0.139	−0.166
	(0.112)	(0.116)
tabloid newspaper	−0.129	-0.326^{**}
	(0.110)	(0.115)
ideological divide: conflict	0.026	0.059
	(0.092)	(0.094)
ideological divide: ambiguous	0.066	0.605^{***}
	(0.151)	(0.152)
Time Bound Factors		
days since start	-0.877^{***}	−0.144
	(0.266)	(0.281)
Year of protest (since 1992)	−0.126	0.194
	(0.187)	(0.193)
AIC	5796.712	5620.873
BIC	5886.397	5710.557
Log Likelihood	−2884.356	−2796.436
Num. obs.	4474	4474
Num. groups: np_year	154	154
Var: np_year (Intercept)	0.190	0.212

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{+}p < 0.1$

Table E.2: Regression Results for Aggregated Frames (without The Times Articles)

	Delegitimising				Legitimising		
	Trouble: police good	Trouble: protesters bad	Decay of morals	Nuisance	Cause: officials bad	Cause: protesters good	Trouble: police bad
(Intercept)	-2.483^{***}	-0.900^{***}	-0.753^{***}	-1.640^{***}	-0.961^{***}	-1.317^{***}	-2.017^{***}
	(0.217)	(0.154)	(0.164)	(0.157)	(0.162)	(0.163)	(0.203)
Event-Level Factors							
goal: anti-war	0.033	-0.061	0.230	0.151	0.037	-0.095	-0.380
	(0.318)	(0.208)	(0.206)	(0.209)	(0.203)	(0.274)	(0.241)
goal: labour protests	-0.077	-0.226^{+}	0.077	0.391^{**}	0.305^{*}	0.122	-0.400^{**}
	(0.194)	(0.132)	(0.122)	(0.130)	(0.127)	(0.157)	(0.151)
goal: police	1.657^{***}	0.649^{*}	-0.935^{**}	-0.878^{*}	-0.157	-0.281	1.849^{***}
	(0.290)	(0.265)	(0.359)	(0.439)	(0.308)	(0.392)	(0.321)
goal: social-issue	-0.488^{**}	-0.132	-0.006	0.059	0.876^{***}	-0.287^{*}	-0.610^{***}
	(0.156)	(0.104)	(0.104)	(0.116)	(0.103)	(0.141)	(0.122)
violent protest	1.278^{***}	0.791^{***}	-0.202^{*}	-0.387^{***}	-0.576^{***}	-0.053	1.545^{***}
	(0.135)	(0.088)	(0.090)	(0.105)	(0.092)	(0.112)	(0.099)
repression of peaceful p.	0.438^{*}	-0.110	-0.472^{***}	0.245^{*}	-0.465^{***}	-0.543^{***}	0.338^{*}
	(0.184)	(0.117)	(0.118)	(0.122)	(0.115)	(0.163)	(0.134)
Outlet-Level Factors							
right-wing	-0.194	-0.006	0.194	0.201	-0.162	-0.367^{**}	-0.161
	(0.181)	(0.129)	(0.139)	(0.131)	(0.135)	(0.142)	(0.173)
tabloid newspaper	0.240	0.145	-0.180	-0.136	-0.218	-0.421^{**}	0.414^{*}
	(0.172)	(0.127)	(0.139)	(0.129)	(0.135)	(0.140)	(0.168)
ideological divide: conflict	0.091	0.184^{+}	-0.034	0.196^{+}	0.082	-0.022	-0.014
	(0.148)	(0.099)	(0.102)	(0.110)	(0.102)	(0.127)	(0.122)
ideological divide: ambiguous	0.680^{**}	0.497^{**}	-0.056	0.262	0.548^{***}	0.305	-0.125
	(0.247)	(0.162)	(0.165)	(0.177)	(0.158)	(0.195)	(0.205)
Time Bound Factors							
days since start	-0.963^{*}	-1.690^{***}	0.461	-0.474	0.124	-1.043^{*}	-0.140
	(0.447)	(0.305)	(0.286)	(0.344)	(0.300)	(0.409)	(0.323)
Year of protest (since 1992)	-0.511^{+}	0.004	-0.297	0.107	0.489^{*}	-0.192	0.812^{**}
	(0.298)	(0.217)	(0.234)	(0.219)	(0.229)	(0.229)	(0.287)
AIC	3084.368	5379.134	5335.349	4205.987	5129.607	3414.299	4622.768
BIC	3174.053	5468.818	5425.034	4295.671	5219.292	3503.984	4712.453
Log Likelihood	-1528.184	-2675.567	-2653.675	-2088.993	-2550.804	-1693.150	-2297.384
Num. obs.	4474	4474	4474	4474	4474	4474	4474
Num. groups: np_year	154	154	154	154	154	154	154
Var: np_year (Intercept)	0.443	0.295	0.376	0.227	0.344	0.190	0.615

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{+}p < 0.1$

Table E.3: Regression Results for Individual Frames (without The Times Articles)

	Delegitimising Frames	Legitimising Frames
(Intercept)	0.354^{***}	-0.345^{***}
	(0.103)	(0.081)
Event-Level Factors		
goal: anti-war	-0.219	-0.024
	(0.145)	(0.157)
goal: labour protests	0.108	0.246^{*}
	(0.098)	(0.098)
goal: police	0.211	-0.217
	(0.238)	(0.247)
goal: social-issue	-0.026	0.624^{***}
	(0.072)	(0.072)
violent protest	0.503^{***}	-0.511^{***}
	(0.066)	(0.066)
repression of peaceful p.	0.157^{+}	-0.619^{***}
	(0.082)	(0.087)
Outlet-Level Factors		
right-wing	0.043	-0.100
	(0.098)	(0.066)
tabloid newspaper	-0.073	-0.278^{***}
	(0.098)	(0.067)
ideological divide: conflict	0.030	-0.160^{*}
	(0.067)	(0.068)
ideological divide: ambiguous	0.233^{+}	0.548^{***}
	(0.121)	(0.120)
Time Bound Factors		
days since start	-0.697^{**}	-0.172
	(0.221)	(0.231)
Year of protest (since 1992)	0.048	0.255^{*}
	(0.109)	(0.108)
AIC	7416.638	7159.245
BIC	7509.562	7252.170
Log Likelihood	-3694.319	-3565.623
Num. obs.	5639	5639
Num. groups: Newspaper	8	8
Var: Newspaper (Intercept)	0.010	0.000

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{+}p < 0.1$

Table E.4: Regression Results for Aggregated Frames (Newspaper Instead of Newspaper per Year as Level 2)

	Delegitimising				Legitimising		
	Trouble: police good	Trouble: protesters bad	Decay of morals	Nuisance	Cause: officials bad	Cause: protesters good	Trouble: police bad
(Intercept)	-2.393^{***}	-0.963^{***}	-0.887^{***}	-1.618^{***}	-0.818^{***}	-1.398^{***}	-1.977^{***}
	(0.149)	(0.125)	(0.103)	(0.102)	(0.089)	(0.135)	(0.111)
Event-Level Factors							
goal: anti-war	−0.161	−0.196	−0.083	0.078	0.110	−0.139	−0.163
	(0.281)	(0.173)	(0.167)	(0.172)	(0.163)	(0.236)	(0.183)
goal: labour protests	0.024	−0.135	0.119	0.416^{***}	0.229^{*}	0.057	-0.411^{***}
	(0.158)	(0.108)	(0.102)	(0.108)	(0.103)	(0.136)	(0.121)
goal: police	1.241^{***}	0.441^{*}	-0.910^{**}	-0.671^{+}	−0.068	−0.383	1.460^{***}
	(0.230)	(0.222)	(0.314)	(0.376)	(0.264)	(0.358)	(0.278)
goal: social-issue	-0.308^{*}	0.152^{*}	0.060	0.101	0.853^{***}	-0.332^{**}	-0.550^{***}
	(0.124)	(0.075)	(0.076)	(0.088)	(0.073)	(0.112)	(0.086)
violent protest	1.434^{***}	0.832^{***}	−0.082	-0.356^{***}	-0.660^{***}	−0.092	1.626^{***}
	(0.110)	(0.068)	(0.069)	(0.085)	(0.071)	(0.091)	(0.074)
repression of peaceful p.	0.450^{**}	−0.102	-0.314^{***}	0.349^{***}	-0.620^{***}	-0.577^{***}	0.168
	(0.156)	(0.095)	(0.092)	(0.096)	(0.090)	(0.137)	(0.103)
Outlet-Level Factors							
right-wing	-0.305^{*}	−0.011	0.161^{+}	0.146^{+}	−0.059	-0.321^{**}	−0.135
	(0.120)	(0.122)	(0.096)	(0.083)	(0.075)	(0.123)	(0.097)
tabloid newspaper	0.090	0.091	−0.069	−0.064	-0.154^{*}	-0.443^{***}	0.330^{***}
	(0.113)	(0.122)	(0.095)	(0.083)	(0.075)	(0.132)	(0.090)
ideological divide: conflict	0.299^{**}	0.171^{*}	-0.156^{*}	0.217^{**}	-0.168^{*}	−0.128	0.178^{*}
	(0.110)	(0.071)	(0.072)	(0.083)	(0.072)	(0.100)	(0.081)
ideological divide: ambiguous	0.640^{**}	0.712^{***}	−0.034	0.411^{**}	0.548^{***}	0.205	−0.108
	(0.208)	(0.124)	(0.126)	(0.139)	(0.121)	(0.159)	(0.160)
Time Bound Factors							
days since start	-1.192^{**}	-1.746^{***}	0.623^{**}	−0.290	0.082	-0.806^{*}	−0.166
	(0.377)	(0.253)	(0.233)	(0.282)	(0.241)	(0.343)	(0.253)
Year of protest (since 1992)	-0.424^{*}	0.175	−0.035	0.023	0.405^{***}	0.190	0.942^{***}
	(0.175)	(0.117)	(0.116)	(0.134)	(0.119)	(0.151)	(0.126)
AIC	3889.225	6908.055	6781.373	5394.509	6606.855	4318.170	6126.156
BIC	3982.149	7000.980	6874.298	5487.434	6699.779	4411.094	6219.081
Log Likelihood	−1930.612	−3440.028	−3376.687	−2683.255	−3289.427	−2145.085	−3049.078
Num. obs.	5639	5639	5639	5639	5639	5639	5639
Num. groups: Newspaper	8	8	8	8	8	8	8
Var: Newspaper (Intercept)	0.005	0.019	0.008	0.000	0.001	0.011	0.005

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{+}p < 0.1$

Table E.5: Regression Results for Individual Frames (Newspaper Instead of Newspaper per Year as Level 2)

	Delegitimising Frames	Legitimising Frames
(Intercept)	0.348^{*}	-0.345^{**}
	(0.140)	(0.111)
Event-Level Factors		
goal: anti-war	-0.215	-0.024
	(0.168)	(0.107)
goal: labour protests	0.109	0.246^{***}
	(0.068)	(0.064)
goal: police	0.223	-0.217
	(0.183)	(0.342)
goal: social-issue	-0.021	0.624^{***}
	(0.058)	(0.061)
violent protest	0.513^{***}	-0.511^{***}
	(0.077)	(0.046)
repression of peaceful p.	0.163^{**}	-0.619^{***}
	(0.062)	(0.100)
Outlet-Level Factors		
right-wing	0.021	-0.100
	(0.095)	(0.116)
tabloid newspaper	-0.059	-0.278^{***}
	(0.079)	(0.071)
ideological divide: conflict	0.031	-0.160
	(0.044)	(0.204)
ideological divide: ambiguous	0.243	0.548^{***}
	(0.180)	(0.129)
Time Bound Factors		
days since start	-0.701^{*}	-0.172
	(0.283)	(0.233)
Year of protest (since 1992)	0.027	0.255
	(0.125)	(0.187)
Num. obs.	5639	5639
Pseudo R^2	0.020	0.069
L.R.	84.620	292.506

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{+}p < 0.1$

Table E.6: Regression Results for Aggregated Frames (Robust Clustered Standard Errors)

	Delegitimising				Legitimising		
	Trouble: police good	Trouble: protesters bad	Decay of morals	Nuisance	Cause: officials bad	Cause: protesters good	Trouble: police bad
(Intercept)	-2.374^{***}	-0.980^{***}	-0.882^{***}	-1.619^{***}	-0.816^{***}	-1.377^{***}	-1.989^{***}
	(0.157)	(0.121)	(0.152)	(0.108)	(0.110)	(0.160)	(0.155)
Event-Level Factors							
goal: anti-war	-0.160	-0.183	-0.088	0.078	0.111	-0.139	-0.155
	(0.173)	(0.125)	(0.177)	(0.293)	(0.129)	(0.158)	(0.252)
goal: labour protests	0.025	-0.135^{+}	0.123	0.416^{***}	0.228^{**}	0.057	-0.413^{***}
	(0.165)	(0.080)	(0.098)	(0.111)	(0.072)	(0.088)	(0.106)
goal: police	1.253^{***}	0.467^{+}	-0.918^{***}	-0.671^{*}	-0.064	-0.398	1.471^{***}
	(0.199)	(0.251)	(0.230)	(0.311)	(0.371)	(0.320)	(0.248)
goal: social-issue	-0.306^{*}	0.160^{**}	0.061	0.101	0.854^{***}	-0.333^{***}	-0.548^{***}
	(0.135)	(0.062)	(0.056)	(0.084)	(0.054)	(0.085)	(0.076)
violent protest	1.436^{***}	0.845^{***}	-0.077	-0.356^{***}	-0.660^{***}	-0.102	1.627^{***}
	(0.117)	(0.049)	(0.066)	(0.067)	(0.076)	(0.125)	(0.114)
repression of peaceful p.	0.451^{**}	-0.094	-0.308^{***}	0.350^{***}	-0.620^{***}	-0.586^{***}	0.167
	(0.159)	(0.128)	(0.079)	(0.093)	(0.121)	(0.099)	(0.136)
Outlet-Level Factors							
right-wing	-0.307^{**}	-0.011	0.124	0.144^{+}	-0.053	-0.325^{**}	-0.114
	(0.118)	(0.118)	(0.105)	(0.082)	(0.143)	(0.103)	(0.221)
tabloid newspaper	0.079	0.131	-0.069	-0.063	-0.153^{*}	-0.472^{***}	0.338^{***}
	(0.117)	(0.112)	(0.099)	(0.066)	(0.070)	(0.128)	(0.082)
ideological divide: conflict	0.297	0.187^{**}	-0.167	0.217^{+}	-0.166	-0.137	0.190
	(0.211)	(0.070)	(0.171)	(0.121)	(0.256)	(0.160)	(0.380)
ideological divide: ambiguous	0.640^{+}	0.749^{***}	-0.043	0.411^{*}	0.550^{***}	0.185^{+}	-0.094
	(0.360)	(0.200)	(0.115)	(0.166)	(0.158)	(0.109)	(0.258)
Time Bound Factors							
days since start	-1.200^{*}	-1.756^{***}	0.622^{*}	-0.291	0.081	-0.805^{*}	-0.165
	(0.542)	(0.294)	(0.306)	(0.221)	(0.239)	(0.356)	(0.430)
Year of protest (since 1992)	-0.452^{*}	0.113	-0.020	0.022	0.395^{*}	0.211	0.932^{***}
	(0.184)	(0.077)	(0.213)	(0.140)	(0.190)	(0.214)	(0.185)
Num. obs.	5639	5639	5639	5639	5639	5639	5639
Pseudo R^2	0.103	0.077	0.011	0.035	0.088	0.034	0.225
L.R.	313.628	320.843	42.856	124.608	362.590	103.886	991.718

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{+}p < 0.1$

Table E.7: Regression Results for Individual Frames (Robust Clustered Standard Errors)