STRUCTURAL CONTINGENCIES AND THE
SOCIAL CONTROL OF PROTEST

A Dissertation in
Sociology
by
Patrick Rafail

© 2012 Patrick Rafail

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2012
The dissertation of Patrick Rafail was reviewed and approved* by the following:

John D. McCarthy
Professor of Sociology
Dissertation Advisor, Chair of Committee

Lee Ann Banaszak
Professor of Political Science

D. Wayne Osgood
Professor of Crime, Law, and Justice and Sociology

Alan Sica
Professor of Sociology

John Iceland
Professor of Sociology and Demography
Head of the Department of Sociology

*Signatures are on file in the Graduate School.
Abstract

The policing of protest is a central component of social movement mobilization. Existing explanations of protest policing in western democratic countries are predicated on rationalist assumptions, which assume that demonstrators’ behavior is the primary factor shaping police response. Yet, such explanations ignore the political and spatial context where protest occurs, as well as the institutional pressures, policies, and preferences that also shape police action. This dissertation challenges the dominance of such rationalist explanations by emphasizing how a wide variety of elements other than protestor behavior strongly condition police responses. I address four major research questions: First, how is protest policing linked to the social structural, institutional, and political-economic environment where it occurs? Second, (how) are elements of a demonstration other than participant behavior related to police responses? Third, has the policing of protest evolved over time? If so, why has it evolved in the way that it did? Finally, do similar factors influence both overt and covert police repression of protest?

To answer these questions I examine three main aspects of protest policing: First, I look at trends in protest policing across 20 major U.S. cities (1996-2006) to see whether there are differences in protest policing by place. The results show extensive variation in police response, even after accounting for the behavior of the protestors. Second, to examine temporal trends in protest policing I analyze several thousand demonstrations occurring in New York City (1960-2006). I find a notable increase in the use of force and arrests at demonstrations over time, which I link to spatial privatization and the adoption of extremely aggressive crime control policies. Finally, to assess whether overt and covert forms of state control follow a similar logic, I examine patterns of covert surveillance of social movement organizations based in Philadelphia (2009-2010). The results indicate a groups’ contentious history had little impact on surveillance rates, which were instead strongly related to groups’ political ideologies.
Overall, I demonstrate that explanations emphasizing behavior are incomplete and perhaps misleading. Furthermore, contextual, temporal, and ideological asymmetries in protest policing represent a form of political inequality and political stratification.
# Table of Contents

List of Figures ......................................................... ix
List of Tables .......................................................... xi
Acknowledgments ......................................................... xii

Chapter 1
Introduction .............................................................. 1
  1.1 Introduction ................................................... 1
  1.2 Major Theories of Protest Policing ............................ 3
    1.2.1 Escalated Force and Negotiated Management .......... 4
    1.2.2 Protest Policing Styles ................................... 5
    1.2.3 Strategic Incapacitation .................................. 6
    1.2.4 Command and Control ..................................... 8
    1.2.5 The Weakness Explanation ................................. 9
    1.2.6 The Threat Explanation .................................... 10
  1.3 Limitations of the Existing Explanations ....................... 11
    1.3.1 The Abundant Resource Bias ............................... 11
    1.3.2 The Reactive Bias ......................................... 12
    1.3.3 The Astructural/Decontextual Bias ....................... 13
  1.4 Plan of This Dissertation ....................................... 14
    1.4.1 Major Research Questions ................................. 14
    1.4.2 Chapter 2: The Impact of Place and Context .......... 15
    1.4.3 Chapter 3: The Temporal Evolution of Protest Control .. 16
    1.4.4 Chapter 4: Patterns of Covert Repression ............. 17
    1.4.5 Appendix A: Statistical Methodology .................... 18
    1.4.6 Appendix B: Cities and Newspaper Sources ............ 19

V
4.2.1.1 Organizational Demands ........................................... 99
4.2.1.2 The Logic of Target Selection ................................. 100
4.2.2 From Subversion to Information Gathering? .................... 101
4.2.2.1 A Brief Overview of COINTELPRO ......................... 102
4.2.2.2 Intelligence Gathering Reformation and 9/11 .............. 103
4.2.2.3 Social Movement Mobilization and Domestic Terrorism ..... 104
4.2.3 Summary and Expectations ........................................ 106
4.3 Data and Methods .................................................... 107
4.3.1 Variables and Measurement ....................................... 111
4.3.2 Analytic Strategy .................................................. 113
4.4 Results .................................................................. 115
4.5 Discussion and Conclusion ........................................... 123

Chapter 5
Conclusion .................................................................. 128
5.1 Synthesizing and Looking Forward ................................. 128
5.2 Major Implications .................................................... 130
5.2.1 Implication 1: Protest Policing Isn't Always Reactive ...... 130
5.2.2 Implication 2: Context Matters ................................. 131
5.2.3 Implication 3: Ideology and Identity Matter ................. 132
5.3 Future Directions ...................................................... 133
5.3.1 Theorizing Space ................................................... 134
5.3.2 Expanding the Theoretical Gaze ............................... 135
5.3.3 Much Ado About Size ............................................. 136
5.3.4 On Crime Control and Protest Control ....................... 137
5.3.5 What About the Police Perspective? ......................... 138
5.4 Conclusion ............................................................. 139

Appendix A
Statistical Methodology .............................................. 141
A.1 Introduction .......................................................... 141
A.2 Guiding Heuristic Principles ...................................... 142
A.2.1 The Box Heuristic ............................................... 143
A.2.2 The Berk-Freedman Heuristic .................................. 144
A.2.3 The Ockham Heuristic .......................................... 145
A.2.4 Summary .......................................................... 148
A.3 The Problems with P-Values ...................................... 148
A.3.1 The Logic of Frequentist Inference ........................... 149
A.3.2 Criticisms of P-Value Based Inference ...................... 152
# List of Figures

2.1 Conceptual model of protest policing ........................................... 33
2.2 The occurrence of arrests at protest events taking place in 20 major U.S. cities, 1996-2006 ($n = 11,426$) .................................................. 42
2.3 Posterior means and 95% credible intervals for situational threats from a Bayesian semiparametric multilevel logistic regression model predicting arrests during protests in 20 U.S. cities, 1996-2006 ($n = 11,426$) .................................................. 45
2.4 Posterior means and 95% credible intervals for protest characteristics from a Bayesian semiparametric multilevel logistic regression model predicting arrests during protests in 20 U.S. cities, 1996-2006 ($n = 11,426$) .................................................. 47
2.5 Posterior means and 95% credible intervals for police tactics from a Bayesian semiparametric multilevel logistic regression model predicting arrests during protests in 20 U.S. cities, 1996-2006 ($n = 11,426$) .................................................. 48
2.6 Posterior means and 95% credible intervals for contextual factors from a Bayesian semiparametric multilevel logistic regression model predicting arrests during protests in 20 U.S. cities, 1996-2006 ($n = 11,426$) .................................................. 49
3.1 Annual property crime and violent crime rates (per 100,000) in New York, NY from the Uniform Crime Reports, 1960-2006 .................. 63
3.2 Percentage of protest events in New York, NY with arrests and the police use of force between 1960 and 2006 ($n=6,147$) ...................... 79
3.3 Percentage of protest events in New York, NY with civil disobedience, property damage, or violence between 1960 and 2006 ($n=6,147$) 80
3.4 Nonparametric smooth of event size from a Bayesian semiparametric multilevel logistic regression model estimating the police use of force at protest events in New York, NY ($n=6,147$) .............. 88
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Movement-level mean deviations in the covert surveillance of Philadelphia social movement organization by the Pennsylvania Office of Homeland Security, 2009-2010 ($n = 431$)</td>
<td>117</td>
</tr>
<tr>
<td>4.2</td>
<td>Posterior means and 95% credible interval from a Bayesian semiparametric logit model predicting the covert surveillance of Philadelphia social movement organizations by the Pennsylvania Office of Homeland Security ($n = 431$)</td>
<td>120</td>
</tr>
<tr>
<td>4.3</td>
<td>Fitted posterior median probabilities and 95% credible intervals of the covert surveillance of Philadelphia social movement organizations by the Pennsylvania Office of Homeland Security across the type of social movement from a Bayesian semiparametric logit model ($n = 431$).</td>
<td>121</td>
</tr>
<tr>
<td>A.1</td>
<td>Simulated nonlinear function and its polynomial approximations</td>
<td>159</td>
</tr>
<tr>
<td>A.2</td>
<td>Nonlinear function with selective heteroskedasticity</td>
<td>161</td>
</tr>
</tbody>
</table>
List of Tables

1.1 Major theoretical explanations of protest policing ............... 4
2.1 Descriptive statistics for predictor variables .................. 43
2.2 Bayesian logistic and multilevel logistic regression posterior means and standard deviations predicting arrests at U.S. protest events, 1996–2006 ................................................. 55
3.1 Key dimensions and differences between the Escalated Force and Negotiated Management models of protest policing ........... 60
3.2 Descriptive statistics for protest events occurring in New York, NY between 1960 and 2006 (n=6,147) ................................. 81
3.3 Bayesian multilevel logistic regression estimates of the factors predicting arrests at protest events in New York, NY between 1960 and 2006 (n=6,147) ............................... 83
3.4 Bayesian semiparametric multilevel logistic regression estimates of the factors predicting police use of force at protest events in New York, NY between 1960 and 2006 (n=6,147) ........................................ 87
4.1 Descriptive statistics for variables used in the analysis (n = 431) ........................................ 116
4.2 Posterior means and standard deviations from a Bayesian semiparametric logistic regression model predicting the surveillance of Philadelphia Social Movement Organizations (n = 431) ..................... 118
A.1 Markov chain iterations for two different starting values ............ 172
B.1 List of Cities and Newspaper Sources ................................ 180
Acknowledgments

Like any scholarly work, this dissertation represents a collective product. I had an exceptionally helpful committee who improved this project in every way and I thank each of them. John McCarthy’s patience, theoretical and methodological insights, and enthusiasm significantly improved this project. He is the sort of scholar that I aspire to be. Alan Sica made just about the best suggestion anyone has ever made to me and I’ll be forever grateful. Our Sunday afternoon chats stand among my favorite graduate school memories. Lee Ann Banaszak’s deep insights into contentious political processes and detailed critiques improved my work on several fronts. Watching Wayne Osgood effortlessly navigate complex technical issues helped me develop my own methodological perspective. The valiant efforts of Sondra Morrison and Melody Lane were also key to completing my degree. Watching them navigate Penn State bureaucracy definitively convinced me that magic is real.

Several research assistants were essential to finish the coding and help refine a few of the variables. They are: Valerie Byer, Amanda Choyce, Emily Duffield, Sarah Kashubsiki, and Courtney Pitman. Their efforts made an impossible project (mostly) possible and I am very grateful to each for their help.

Along the way, I had the best circle of friends that a person could ever ask for. I am lucky to know Hrag, Jenn, John, Rich, Claudia, Katie D., Sara, Tanya, Sean, Andy, Dave, Trög, Sadie, Basil, Pete, and the countless others I have forgotten. Above all else, Katie was there through the worst of times and the worst of times. I couldn’t have done this without her. My wonderful parents, Ehab and Judy, were also extremely supportive over my entire education and I can’t thank them enough.

Finally, I am forever grateful to the hackers, programmers, and participants who have contributed to the development of free software. In particular, the amazing folks behind the R project, JAGS, BUGS, BayesX, Emacs, ESS, AUCTeX, GNU/Linux, GCC, Octave, QGIS, Libre Office, ETeX, Perl, and Python deserve a continual round of applause. I cannot overstate the time and effort these pro-
grams saved me relative to the alienating, bloated, paternalistic, and profoundly anti-scientific world of proprietary software.

Pieces of this dissertation were supported by the National Science Foundation (SES-0927875), the Social Sciences and Humanities Research Council of Canada (#742-2006-2135), and a RGSO Dissertation Support Grant from the College of the Liberal Arts at Penn State. The opinions expressed here are my own, and need not reflect the views of any of these institutions.
Introduction

1.1 Introduction

Reflecting on his travels to the United States in 1831, Alexis De Tocqueville noted that the voluntary associations which permeated American society played a key role in preventing the despotic rule of political parties (De Tocqueville, 2004 [1835], 219). Despite arguments by some that American associational life has declined significantly in recent years (e.g. Putnam, 2001), participation in social movement organizations and activism has remained a vibrant part of the political landscape. Individual involvement in social movement activism has experienced its own ebb and flow of activity, particularly since the flurry of protests in the 1960s, yet organized dissent has been and remains to be a major outlet of associational life that can have major societal consequences. As I write this, for example, both the Tea Party and Occupy Wall Street movements have shifted American political discourse, though it remains to be seen whether this will translate to legislative outcomes or if we are witnessing the beginning of a new wave of progressive and conservative activism (see Skocpol and Williamson, 2012 and McCarthy et al., Forthcoming for discussions of the Tea Party and Occupy Wall Street movements
respectively).

A key aspect of social movement activism involves accessible social space: the space for participants to gather, to mobilize and organize without suppression by the state or private actors, and to articulate claims in a public forum. Indeed the right to engage in social movement mobilization is generally recognized to be a cornerstone of democratic society, and is codified in documents ranging from the U.S. Constitution to the United Nations’ Universal Declaration of Human Rights. This dissertation examines a small pocket of social movement mobilization, but one that has dramatic implications for gauging the health of the civil society de Tocqueville found to be so important. In the simplest sense, I examine how and why the state may use its power coercively during the course of protest events, which has come to be referred to policing of protest. I follow della Porta’s (1995) standard definition of protest policing, who argues it is the “police handling of protest events” (p. 55; see also della Porta and Fillieule, 2004).

Many of the existing explanations of variations in protest policing are either explicitly or implicitly rationalist in character (Cunningham, 2003a). That is, they are grounded in what Davenport (2007a) calls the “law of coercive responsiveness,” where the state use of force is proportionate to the threat posed by the mobilizing group. At first glance, there appears to be ubiquitous support for this explanation, which is neatly justified since only peaceful forms of political activism are legally protected. Yet there is a notable and growing body of literature suggests that the social space required for political protest has begun to steadily and stealthily erode (a review is provided by Zick, 2009). And, increasingly heavy-handed protest control has become a core component of how the police respond to social movement activism in western democratic states (Vitale, 2005, 2007; Rafail, 2010; Starr et al., 2011).
This project provides a critique of the rationalist model of state repression and is intended to start a conversation among scholars that expands our theoretical understanding of protest policing beyond rationalist reductionism. The core argument throughout this dissertation is that the non-behavioral aspects of protest and social movement mobilization are fundamental to understanding the policing of protest. To demonstrate this I point to linkages between the literature on contentious politics and other topics that have hitherto been unexamined: policing and crime control, social control, social stratification. “Non-behavioral aspects” is used in a broad sense that is intended to cover not only the social structural context of where a protest occurs but also factors that are completely legal, but may nonetheless alter police responses (e.g. the ideology or demographic composition of the protestors). To proceed, this introductory chapter first reviews the major explanations of protest policing, followed by a brief critique of the existing literature, and then outlines the plan of my dissertation.

1.2 Major Theories of Protest Policing

The literature on the policing of protest, and state repression more generally, is vast and has grown extensively over the last 20 years. As a result I do not attempt to provide an exhaustive literature review of this prohibitively large literature; the most recent reviews are provided by Davenport (2007a) and Earl (2011). Instead, I more selectively focus on the major theoretical models that are commonly used to explain protest policing. Since I am primarily concerned with the control of political protests, I omit more general discussions of collective violence, such as the flashpoint model (Waddington et al., 1989) or McPhail’s (1991) discussion of collective behavior, as well as topics such as the policing of disorderly student
gatherings (McCarthy et al., 2007; Martin et al., 2009).

To facilitate the discussion I distinguish between two major axes of that can be used to differentiate theories of protest policing. First, I distinguish between the models that are and are not *ideal typical*. That is, one subset of protest policing theory is based on constructing broad models intended to capture specific styles of protest policing that form the basis of police conduct. Explanations that are not ideal typical, in contrast, assign analytic primacy to a small number of explanatory variables that are directly observable, and I accordingly refer to these as *mechanistic* explanations. Second, I distinguish between theories that are or are not intended to *generalize* to a broad range of protests and locations. Table 1.1 contains six major accounts of protest policing as well as their explanatory scope across each dimension.

<table>
<thead>
<tr>
<th>Major theoretical explanations of protest policing</th>
<th>Scope of Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ideal Typical</td>
</tr>
<tr>
<td>Escalated Force &amp; Negotiated Management</td>
<td>Yes</td>
</tr>
<tr>
<td>Protest Policing Styles</td>
<td>Yes</td>
</tr>
<tr>
<td>Strategic Incapacitation</td>
<td>Yes</td>
</tr>
<tr>
<td>Command and Control</td>
<td>Yes</td>
</tr>
<tr>
<td>The Weakness Explanation</td>
<td>No</td>
</tr>
<tr>
<td>The Threat Explanation</td>
<td>No</td>
</tr>
</tbody>
</table>

### 1.2.1 Escalated Force and Negotiated Management

McPhail et al. (1998) provide the authoritative statement of the escalated force and negotiated management models of protest policing. As an ideal typical explanation, the authors construct broad styles of protest policing intended to generalize to the majority of protest events occurring in particular historical periods. Under the escalated force style of policing, law enforcement relies primarily on violence,
arrests, and other forms of coercion when engaging with demonstrators. This style of protest control was the standard police response for demonstrations occurring between 1960 and the end of the 1970s. In contrast, the negotiated management style came to dominate in the 1980s and remained the standard police strategy to the present. The logic of negotiated management is based on mutual cooperation and communication between law enforcement agents and demonstrators, where protest is recognized as a form of political expression that should be protected. Unlike the escalated force model, the negotiated management style is explicitly intended to reduce potential conflict, though the police will not hesitate to use force or make arrests if the conduct of the protestors threatens public safety. While McPhail and colleagues acknowledge that the police may occasionally deviate from these overall patterns during their respective historical period, they do provide a set of both general and specific hypotheses that collectively yield a baseline to evaluate the relative permissiveness of protest policing over time. The transition from escalated force to negotiated management is a key component of Chapter 3 (see Section 3.2.1), where it is discussed more thoroughly.

1.2.2 Protest Policing Styles

Donatella della Porta provides an account of state repression that emphasizes protest policing styles (see della Porta, 1995; della Porta and Reiter, 1998; della Porta and Fillieule, 2004), which is intended to be both ideal typical and generalizable to a variety of western democratic states. To conceptualize ‘styles,’ della Porta develops a series of contrasts in police behavior. When aggregated, these contrasts can be used to categorize the relative severity or permissiveness of protest policing. della Porta and Reiter (1998, p. 6) provide several examples the contrasts
in police behavior, such as:

- “brutal” vs. “soft”: The degree of force used by the police during an event
- repressive vs. tolerant: The number of actions that the police prohibit and the degree to which they adhere to existing regulations
- confrontational vs. consensual: The degree to which the police are willing to communicate with the protestors
- reactive vs. preventive: If the protestors engage in illegal activities does law enforcement react immediately or do they confront the demonstrators proactively

The general pattern of police conduct as well as the institutional arrangements of the police form distinct styles of protest policing (though a concrete typology is not provided). In more recent work, (della Porta and Fillieule, 2004) suggest that by the 1990s the police have developed a style emphasizing tolerance and negotiation, though there is recognition that they can be quite antagonistic when dealing with specific groups. As well, Della Porta et al.’s (2006) work on the police of transnational protest events points to a shift away from negotiation and towards coercive control.

### 1.2.3 Strategic Incapacitation

The strategic incapacitation model was proposed in a series of studies by Noakes et al. (2005), Noakes and Gillham (2006, 2007), and Gillham and Noakes (2007). Similar to the escalated force/negotiated management dichotomy the strategic incapacitation hypothesis refers to an ideal typical style of protest policing. However, it is difficult to generalize the model beyond a small and specific subset of protest events (see Section 1.3.1 below for discussion). Noakes, Gillham, and their
collaborators draw from Tilly’s (2000) distinction between contained contention and transgressive contention. Contained contention refers to protest events where all involved actors follow widely established and understood patterns of behavior (i.e. a standard peaceful protest event). When dealing with contained contentious events, the police are much more likely to adopt tactics consistent with a negotiated management strategy. Transgressive contention is more volatile and occurs when protestors use innovative or otherwise unpredictable tactics that alter how and whether the authorities can control the space of the protest site. During transgressive protests the police are much more likely to rely on an aggressive approach to crime control: the strategic incapacitation model.

Noakes et al. (2005, 241) identify five major components of the strategic incapacitation approach, each of which has an explicitly spatial component. First, the police routinely establish zones where protest activity is prohibited using both the built environment of the protest site as well as complex structures of barricades and fencing. Second, there is a disregard for the “safe spaces” of a protest event where activists congregate to eat, sleep, or plan subsequent actions, which are often raided by the police. Third, law enforcement officers make more use of less-lethal technologies that results in more extensive control over the protest site while minimizing major injuries.¹ Fourth, the police use advances in electronic surveillance technology coupled with constant information sharing to monitor the protestors and protest site prior to and during the event. Fifth and finally, the police rely on preemptive and targeted arrests to place leaders in custody combined with mass

¹ Less-lethal weapons represent a broad class of technologies that are intended to stun or maim rather than kill (though the possibility of death is not ruled out). Examples include tear gas, pepper spray, rubber bullets, knee-knockers, beanbag rounds. See McPhail and McCarthy (2005) for further discussion of how these tools came to be used in crowd control situations and Kraska and Kappeler (1997) or Kraska and Paulsen (1997) on the routinization of paramilitary policing.
arrests of mainly peaceful participants in hopes of creating confusion among the
protestors and demobilizing the event as whole.

1.2.4 Command and Control

The command and control model arises from Alex Vitale’s ethnographic research
on how the New York Police Department manages major protest events (see Vitale,
extremely coercive approach to protest control, however, Vitale explicitly limits its
scope to major protest events in New York. Similar to the strategic incapacitation
explanation, command and control policing is based on a micromanaging of the
space where a protest event occurs. Following the Broken Windows\(^2\) philosophy of
policing, command and control protest policing has a zero-tolerance approach to
even the slightest legal transgressions.

Vitale (2005) outlines five major components of the command and control
model. First, the police have a strong aversion to any form of disruption, and
therefore take any necessary steps to ensure that a demonstration does not inhibit
the normal functioning of a protest site. Second, the police use physical structures
such as barricades to not only separate protestors from the public, but also to
retain control over all points of entry or exit to a demonstration. Third, barricades
are also a component of a ‘divide-and-conquer’ strategy where the police cordon
the demonstrators into smaller more manageable groups. This allows the police
relatively unfettered access to any part of a demonstration while also serving as
another way limit to protestor mobility. Fourth, the police use ‘shock and awe’

\(^2\) Broken Windows is a philosophy of policing and crime control outlined in Wilson and Kelling
(1982) and Kelling and Coles (1996) where law enforcement adopts aggressive zero-tolerance
policies to any behavior viewed as disruptive to public order. The Broken Windows perspective
is discussed much more extensively in Chapter 2 (Section 2.1.2.1) and Chapter 3 (Section 3.2.3.2).
tactics during crowd control situations, where a large number of police officers are present or extensive displays of force are used to control or intimidate protestors. Last, as mentioned above, the police adopt a zero-tolerance policy where all laws, municipal regulations, and parameters of the protest permit are adhered to quite closely, though there remains a degree of police discretion.

1.2.5 The Weakness Explanation

The weakness explanation is generally linked to work by Gamson (1975) and focuses on how the characteristics of the challenging group might influence state responses. As noted in Table 1.1, the weakness explanation is mechanistic rather than ideal typical, but it is intended to generalize to a wide variety of protest events. The central thesis is that government actors are hesitant to use coercion against strong actors who may have the resources to resist repressive attempts. Instead, state repression is targeted towards weaker actors since the likelihood that repressive campaigns will succeed is much higher. Studies explicitly advocating a weakness explanation have become rare in contemporary analysis, though it is often invoked as a strawman argument. There is a generally a lack of supporting evidence for the weakness account (Earl et al., 2003; Earl, 2006), but this may be more due to an ambiguity in differentiating between ‘strong’ and ‘weak’ actors and difficulty in linking vague assessments of the relative strength of a challenging group to the behavior of the authorities during a specific protest event. Nonetheless the weakness explanation is useful as it provides a testable hypothesis about how the state’s perception of a protest rather than the behavior of the participants might shape police responses.
1.2.6 The Threat Explanation

The threat explanation is among the most widely invoked models to account for protest policing and has a good deal of empirical evidence supporting its basic tenants (Earl et al., 2003; Soule and Davenport, 2009). The threat explanation is mechanistic rather than ideal typical and is intended to generalize to all protests occurring in western democratic states. At present its strongest advocates are Jennifer Earl (Earl et al., 2003; Earl, 2006, 2011) and Christian Davenport (Davenport, 2000, 2007a,b). The crux of the threat explanation is the intuitive expectation that when a protest becomes threatening (broadly defined) agents of the state are likely to resort to force. According to Davenport (2007a) the amount of force doled out by the state is proportionately tempered by the degree to which an event is threatening, though there are clear boundaries (e.g. deadly force) that the police will not cross except in the rarest of situations (McCarthy and McPhail, 1998).

A major concern with the threat explanation is that its central concept, threat, is ambiguously defined. That is, it is often unclear who is being threatened by contentious political action and how threat concretely translates into police behavior. Early forms of the threat explanation focused on how social movement activism might threaten political and economic elites, but did not account for the preferences of the police, who have their own set of priorities that need not be consistent with those held by elite actors (Earl, 2006). The mechanisms through which elites—which were nearly always ambiguously defined—were able to concretely influence police actions during protest events remained unaddressed almost entirely.

Research by Earl and Soule (2006) has been instrumental in clarifying the concept of threat. They propose a blue-centered model of protest policing intended to shift the focus away from nebulous hypotheses about elite actors to how the police
perceive a demonstration. The blue-centered model emphasizes the immediate situational threats to police safety or public order that can occur during the course of a demonstration, and how such threats are primarily responsible for shaping police responses. Cast in terms of situational threats, Soule and Davenport (2009) find that the behaviors of the protestors are consistent predictors of police action at U.S. events occurring between 1960 and 1990. Examining the impact of contentious behaviors and tactics used by protestors currently provides the central operationalization of the threat explanation.

1.3 Limitations of the Existing Explanations

The current literature on protest policing provides an excellent starting point to make sense of why the state generally allocates coercion to some protests and not others. As it currently stands, however, there are major limitations to the explanatory power of each approach to protest policing. In what follows, I identify three theoretical gaps in the models of protest control described above. A more extensive critique of the various theories is outlined in the chapters below, yet it is also useful to outline and name some general themes to situate those arguments.

1.3.1 The Abundant Resource Bias

There is a great deal of evidence supporting the central hypotheses of both command and control and strategic incapacitation explanations, particularly for major protest events targeting global trade (Ericson and Doyle, 1999; Della Porta et al., 2006). Such events are exceedingly rare, however, and the limited explanatory scope of such models significantly reduces their theoretical utility for the majority of protests. Research examining the security costs for global justice demonstra-
tions by Starr et al. (2011) points towards a dramatic increase since the World Trade Organization meetings in Seattle held in 1999. Costs now range in the hundreds of millions, and the recent G8/G20 summits in Toronto reached nearly a billion Canadian dollars. In other words, as I have noted elsewhere (Rafail, 2010), the amount of resources needed to control major events—including capital, labor power, police coordination—is prohibitive. Specific theoretical explanations of how the police mobilize for major events are quite useful, but the exceptional circumstances and planning accompanying large demonstrations makes it effectively impossible for these protest policing strategies to become regularly implemented. Instead, it is important to develop a theoretical model sufficiently flexible to not only account for large-scale police mobilizations at global justice protests, but a sparsely attended candlelight vigil as well.

1.3.2 The Reactive Bias

A central objection to the threat explanation is its implicit assumptions. Unlike the other major theories of protest policing which explicitly acknowledge that non-behavioral factors might influence the police’s decision to use coercive force—and more importantly, that such decisions need not reflect the rule of law—the threat perspective is strongly predicated on reactive police conduct. The key variables used to evaluate the empirical support for the threat explanation includes contentious tactics (e.g. civil disobedience, property damage) or event size (see for example Earl et al., 2003; Earl and Soule, 2006; Soule and Davenport, 2009). A long line of research on policing, including classic works by Bittner (1967, 1970) and Manning (1977), points towards the central role of proactive policing, which is more difficult to reconcile with the threat explanation. In the context of protest
control, it is therefore reasonable to expect that the police do not simply (or perhaps primarily) react to tactics that threaten. Instead, a variety of non-behavior aspects of a protest ranging from its spatial location to the claims of the protestors that fall outside the purview of behavioral threats are also significant and might also reshape police perceptions of whether behavior is threatening.

1.3.3 The Astructural/Decontextual Bias

A third limitation relates to the lack of structural awareness or social context in many explanations of protest policing. Some research has outlined how existing political opportunities can affect protest control (della Porta, 1995), however, less attention has been paid to how existing structural and institutional arrangements might inform police conduct. These comments do not apply universally; McPhail et al.'s (1998) analysis of protest policing over time does draw on a series of legal decisions and the evolution of law enforcement policies. Much of the remaining literature treats structural arrangements as a given and instead emphasizes characteristics of the challengers or their behavior. This results in an astructural movement-centric bias in protest policing theory that is largely uninformed by broader patterns in social control.

It is difficult to imagine that social movement mobilization occurs in isolation from other societal processes. Opening theoretical space for social structural and institutional contexts could provide hitherto unexamined insights on how and why repression occurs. Recent work by Oliver (2008) provides an excellent example. She demonstrates that the mass incarceration of African Americans is a central factor explaining the decline of the civil rights movements and black activism. Punitive systems are far outside the scope of the theoretical models described
above, yet in the context described by Oliver, they provide an intuitive and powerful way to understand movement decline. In short, much more research is needed to examine how protest policing is linked to crime control, social control, and social stratification systems.

1.4 Plan of This Dissertation

This dissertation contains three substantive chapters each addressing a different aspect of protest policing. There are also two appendices: Appendix A discusses the technical details of my analytic strategy and Appendix B lists the cities and newspapers sources used to construct the data in the statistical analysis in Chapter 2, Chapter 3, and Chapter 4. I now turn to an outline of my major research questions and then provide a brief discussion of each substantive chapter and the first appendix.

1.4.1 Major Research Questions

The discussion above has reviewed the major theories of protest policing and outlined some limitations of these accounts. To begin to develop a more comprehensive explanation of protest policing policing, this dissertation examines four major research questions that are intended to expand and improve on scholarly understandings of how protest is managed and controlled. The central questions I pose in this project are:

1. How is protest policing linked to the social structural, institutional, and political-economic environment where it occurs?

2. Are elements of a demonstration other than the behavior of the participants
systematically related to police responses? If so, in what way?

3. Has the policing of protest evolved over time? If so, why has it evolved in the way that it did?

4. Do the factors that influence overt police repression of protest carry over to covert repression as well?

These research questions provide the theoretical motivation for me to significantly extend current scholarly understanding of protest policing. To answer these questions requires a careful examination of the contextual, temporal, structural bases of protest policing, which has only received little attention.

1.4.2 Chapter 2: The Impact of Place and Context

Chapter 2 begins my critique of the astructural bias common in explanations of protest policing. I suggest that existing scholarship on protest control has predominantly focused on the effect of situational threats such as civil disobedience or violence on police action. This has resulted in a theoretical understanding of protest control that is largely autonomous from broader trends in formal state-sponsored social control. Using variation in protest policing across large American cities, I demonstrate that a focus on smaller analytic units, particularly cities, can provide a viable way to link protest policing to social structural factors. I particularly emphasize institutional change in police departments, shifting police training regimens and strategies, and the widely established link between political conservatism and punitiveness. Using a sample of 11,426 protest events in 20 major U.S. cities, I estimate a series of Bayesian semiparametric mixed logistic regression models predicting arrests. The results suggest that even after accounting
for the behaviors of the protestors, there remains considerable city-level variability in protest policing. Furthermore, the results suggest that more liberal cities are more repressive, that protest in wealthier cities is comparatively privileged, and that the adoption of Broken Windows policing strategies in many major cities may result in more aggressive protest control strategies.

1.4.3 Chapter 3: The Temporal Evolution of Protest Control

In Chapter 3 I more directly examine temporal trends in the policing of protest since the wave of contention beginning in the 1960s. Scholars have suggested that the policing of protest has become less aggressive since wave of contention starting in the 1960s. While widely accepted, studies affirming a softening of police conduct have focused on national trends despite awareness of an uneven diffusion of police tactics across the United States. This chapter examines the temporal trends in New York, NY between 1960 and 2006 to evaluate whether, how, and why dominant strategies of protest policing have changed over time. Drawing on the widespread privatization of public space in New York during the 1980s, coupled with the adoption of Broken Windows crime control strategies, I develop the concept of policy spillover, which provides an alternative explanation of how a series of municipal policy decisions had unintentional spillover effects resulting in the adoption of more aggressive protest policing strategies. Using Bayesian multilevel logistic regression models and a sample of 6,147 protest events occurring in New York, NY between 1960 and 2006, I confirm that the prevalence of arrests and other forms of police force have actually increased over time even though illegal or other highly contentious tactics have declined. The results additionally suggest
that there is not a direct relationship between crime rates and either arrests or the use of force. This indicates that the political construction and control of crime may be more important than the absolute amount of crime when it comes to protest policing, which is consistent with the proposed policy spillover explanation.

1.4.4 Chapter 4: Patterns of Covert Repression

While Chapters 2 and 3 examine overt forms of protest policing, Chapter 4 expands the type of police behavior I examine by looking at trends in covert surveillance. Scholarship on the state control of social movements has predominately focused on overt forms of repression. This has resulted in relatively little attention to trends in more covert forms of control. There have been recent anecdotal suggestions by researchers that government surveillance of social movement organizations (SMOs) has become widespread and routinized after the 9/11 terrorist attacks. Unlike previous forms of government surveillance and infiltration, which were aimed at disrupting or disbanding a group deemed threatening or radical by the state, contemporary surveillance is grounded in a logic of information gathering that has diffused across law enforcement agencies since the 9/11 attacks. As a result, government actors now cast a large net and surveil a wide variety of groups. A consequence of this is that traditional factors predicting surveillance, such as staging contentious protests, are no longer the major determinant of covert surveillance. Instead, organizations become threatening based on their ideology. To test this explanation and determine the major predictors of covert surveillance, I analyze a database of 431 Philadelphia SMOs active between January, 1996 and October, 2009. This is linked to a list of organizations surveilled by the Pennsylvania Office of Homeland Security (PA-OHS) between November, 2009 and September, 2010.
Results from a Bayesian semiparametric logistic regression analysis suggest that there is no meaningful relationship between the use of contentious or violent protest tactic and surveillance by PA-OHS. Instead, covert surveillance was strongly linked to the ideology of an organization. This is particularly true of groups focusing on peace, the environment, animal rights, or other progressive causes. The results also point towards a complex, non-linear, relationship between media attention and the probability that a group was surveilled.

1.4.5 Appendix A: Statistical Methodology

The central analytic strategy I use in this project—Bayesian semi-parametric generalized mixed regression—is rarely used in empirical social research. As the use of increasingly complex statistical models becomes commonplace there has not been a corresponding discussion critically examining the implications of such analyses. There is a vast and growing literature suggesting that widely implemented techniques are routinely misused (see e.g. Freedman, 2006, on the incorrect use of robust standard errors) and pointing to deep problems in the mischaracterization of frequentist inference (Ziliak and McCloskey, 2008). Instead, quantitative analysis is assumed to be problem-free (Agger, 2007), and assumptions that are (at best) quite questionable are granted without worry. The purpose of Appendix A is, on the one hand, to provided a critique of several problematic practices common in empirical sociology, as well as to flush out the epistemological and methodological details of my analytic strategy in some detail.
1.4.6 Appendix B: Cities and Newspaper Sources

The final appendix contains a table outlining the cities and newspaper sources used in each of the chapters.

From the “battle of Seattle” to a small candlelight vigil, the factors influencing police responses to protest are an integral component to a holistic theory of social movement mobilization. A significant portion of the literature on protest policing, including a series of studies by della Porta (1995), McPhail et al. (1998), and others (Noakes et al., 2005; Noakes and Gillham, 2006) have proposed ideal typical models of protest policing that hypothesize patterns of police action, typically of all police within a nation-state. While such models of protest policing are quite useful to generate expectations about police conduct, they have the potential to overgeneralize the uniformity of protest control and its predictors. With few exceptions (see Earl and Soule, 2006; Oliver, 2008), much of the research on protest policing has focused on the movement-side dynamics influencing protest control. That is, conceptualizations of repression have focused on movement related fac-
tors such as contentious behaviors, but have paid comparatively little attention to placing protest policing within the larger context of state-sponsored social control.

This lack of attention is surprising, since it is well established that the organizational structure of law enforcement is highly decentralized and quite variable in many western democratic states (Bayley, 1992). As a result, there are important questions of how focusing on national trends obfuscates theoretically meaningful variation within and between nation-states. A growing number of studies—for example Wisler and Kriesi (1998), McAdam et al. (2005), Sampson et al. (2005), Vitale (2005, 2007), and Rafail (2010)—suggest that the prevailing emphasis on movement-side dynamics and national trends level may have created a theoretical blind-spot in existing scholarship, making it difficult to integrate broader changes in law enforcement and social control into a comprehensive explanation of protest policing.

This study capitalizes on variation in protest policing across major American cities to link protest control to state control. My core argument is that variation in protest policing across municipal police departments can be used to develop mechanistic pathways linking meso-structural social, economic, and political contexts to protest policing. To support this contention, this research provides among the first large-scale analyses of whether the conclusions on protest-policing dynamics drawn from nation-level studies hold once they are examined at the city level. Using a sample of over 11,000 protest events taking place in 20 major American cities between 1996 and 2006, I demonstrate that not only is there considerable variation in protest policing across the United States, but that this variability is closely related to city-level contextual and structural factors.
2.1 State Repression and Protest Policing

State repression refers to state actions that violate human rights or dignity as well as acts of violence or intimidation (Stockdill, 2003). More specifically, Davenport (2007a) argues that state repression is the use of state power to violate First Amendment-type rights, which include freedom of speech, assembly, travel, association, belief, and peaceful dissent (p. 2). Scholars generally focus on the negative sanctions applied by states, and purposively exclude a wide variety of behaviors ranging from crime control to the use of positive sanctions. State repression includes a variety of behaviors ranging from egregious violations of human rights (e.g. genocide) to the repression of social movements. Protest policing, or the police handling of protest events (della Porta, 1995, p. 55), is a subset of state repression that has generally focused on overt police responses to social movement mobilization (see e.g. McPhail et al., 1998; Earl et al., 2003; Earl and Soule, 2006; Soule and Davenport, 2009).

In western democratic countries, the policing of protest is grounded in public order management strategies (POMS). According to McCarthy and McPhail (1998, p. 91), POMS are “...the more or less elaborated, more or less permanent organizational forms, their guiding policies and programs, technologies, and standard policing practices that are designed by authorities for supervising protesters’ access to public space and managing them in that space.” Rafail (Forthcoming) argues that contemporary POMS emphasize spatial and temporal control over the sites where protest occurs coupled with a discursive preference on negotiating with protest groups, though this preference is not always practiced by law enforcement. Many of the key theoretical questions on protest policing center around the factors that influence POMS.
One of the most robust findings in the literature on protest control is that when protesters use violent or other highly contentious tactics that threaten the police or political elites, governments will generally respond repressively (Davenport, 2000; Earl et al., 2003; Earl and Soule, 2006; Soule and Davenport, 2009). The rationale for this relationship is quite intuitive: threatening events may challenge elite dominance, police authority, or disrupt public order, and repression allows the state or its agents to reassert their dominance. The threat-repression relationship is regarded as sufficiently important that Davenport (2007a) argues it is law-like, and refers to such reactive state repression as the “law of coercive responsiveness” (p. 8).

Despite its intuitive appeal, however, research by Earl (2006) has problematized the coupling of threats to elites and threats to the police. She argues that linking the demands of elite actors to the behavior of the police requires several key assumptions about the factors influencing police conduct, including how the desires of the elites translate to the officers on the ground. Building on this line of inquiry, Earl and Soule (2006) examine the notion of threat and propose a blue-centered approach to understanding POMS. They begin by differentiating between threats to elites and threats to the police officers responsible for controlling protest events. After examining police responses to protest events in New York State between 1968 and 1973, Earl and Soule (2006) conclude that threats to elites have limited explanatory power in explaining protest policing. Instead, situational threats, which refer to behavioral challenges to the police control over a demonstration, are of central importance to understanding police presence and action. Based on these results, Earl and Soule (2006) argue that threat must be grounded in police perceptions of protestor conduct rather than relatively vague connections between elite actors influencing police behaviors.
2.1.1 Limitations of Existing Approaches to Protest Policing

Even though the threat-based approaches to protest policing have relatively high explanatory power, a growing body of literature suggests that the primary emphasis on threat may oversimplify the repression of social movements. Oliver (2008) has criticized existing conceptualizations of state repression more broadly due to a relatively strict focus on movement-side dynamics of protest policing. She notes that standard definitions of state repression, such as Davenport's (2007a) discussed above, largely ignore the linkages between the control of social movements and other forms of state control. Focusing on the incarceration of African Americans in the United States, Oliver convincingly demonstrates that patterns of crime control are a form of state repression that considerably influence the mobilization patterns of socially marginalized groups. Additionally, recent efforts by Earl (2003, 2004) have laid the groundwork towards a more exhaustive and flexible conceptualization of repression. Earl differentiates between the identity of the repressive actor, the type of repressive activity, and whether repressive actions are observable or unobservable (see Earl, 2003, Table 2), presenting a series of hypotheses for future research on repression.

Both Oliver (2008) and Earl (2003, 2004) point to a broader problem apparent in scholarship on protest policing: It is not simply that protest policing has been conceptualized too narrowly but that social movement scholars have constructed theoretical models of protest policing that are largely autonomous from and resilient to broader trends in law enforcement and social control widely noted by political sociologists and sociologists of punishment. For example, innovation in crime control practices—whether in terms of technological advancement, more effec-
tive training procedures, or other developments—may have spillover effects on how the police approach protest control. Ignoring such changes risks misinterpreting the underlying causal factors that shape how the police respond to protest events. It follows that careful attention to the prevailing trends in crime control is quite important, even if one is unwilling to widen the scope of what constitutes repressive behavior as Oliver (2008) suggests. In the next section, I expand on this argument in more detail, suggesting that shifting our focus from national-level protest policing patterns to cities can provide a feasible way to link protest policing to social control.

### 2.1.2 Linking Protest Policing to Broader Systems of Social Control

As mentioned above, a considerable drawback to much scholarship cited above is a problem of scope. A substantial portion of the literature focuses on the mechanisms related to social movement processes for entire nation states (see for instance McCarthy and McPhail, 1998; Soule and Davenport, 2009) or comparatively smaller units such as U.S. states (e.g. Earl et al., 2003; Earl and Soule, 2006). This has been a major obstacle to the identification of mechanisms through which protest policing is linked to social control more broadly. I propose that examining variability in protest policing in more fine-grained units of analysis, particularly cities, can provide mechanistic pathways through which we can identify which structural and institutional features influence protest control.

Scholars have generally acknowledged that there is considerable variation in protest policing. For instance, in their analysis of American protest policing between 1960 and 1995, McPhail et al. (1998, p. 68) conclude that the adoption
of any particular POMS is far from uniform across American municipalities, and recommend that future research pays more attention to such dynamics. Likewise, della Porta and Reiter (1998) argue that the police have long institutional memories, and that previous interactions with social movement actors can have a profound effect on subsequent events. Few studies have directly engaged with the factors shaping the institutional memory of the police, and its geographic scope in particular. A focus on national trends in protest control implicitly assumes that the police’s institutional memory and preferred protest management styles can be shaped by events occurring in distant locations, or that the behavior of the police is roughly equivalent across a country. Or, at least, that any observed variance is not theoretically interesting. While it is almost certainly the case that events such as the World Trade Organization meetings in Seattle or other major protest may have national or transnational repercussions, recent work by McAdam et al. (2005) suggests that social movement mobilization has become increasingly local in scope. As a result, it is likely that much of the knowledge informing protest policing stems from local contention.

While geographic units such as counties provide alternatives to focusing on cities, I suggest that there are several advantages to cities. The geographic boundaries of a city generally coincide with the jurisdiction of its municipal police department, making cities a good match to the law enforcement organization(s) responsible for the majority of protest control. This allows the possibility to examine and compare the effect of institutional and organizational features of different police departments. As well, due to their high population density, a majority of protest takes place in major cities. This is not to say that all protest occurs in major cities; research has suggested that there is considerable mobilization in suburbs around major cities (McAdam et al., 2005). Unlike suburbs, however, cities do
provide relatively contained geographic units with a more manageable number of
police departments. In what follows, I propose three major mechanisms linking
protest control to broader systems of state repression. The goal is not to pro-
vide an exhaustive enumeration of the myriad factors that could influence protest
policing, but instead to focus on major factors that are known to have consider-
able effects on law enforcement, and could plausibly influence protest control as
well. These include (1) institutional changes in law enforcement; (2) the evolution
of police training and tactics for public order policing, and (3) the link between
conservatism and higher levels of punitiveness.

2.1.2.1 Institutional Change in Law Enforcement

Among the most notable institutional changes in police departments in several
major municipalities is the adoption of a broken windows policing philosophy. The
seminal statement of this philosophy comes from Wilson and Kelling (1982) who
link criminal behavior with the degree to which a neighborhood or community
appears to be in a state of disrepair. Areas that have unrepaired broken windows
or high levels of graffiti signal a lack of formal and informal social control to
criminal actors. To reduce the rate of more serious crimes such as homicide or
sexual violence, Wilson and Kelling (1982) and other advocates of broken windows
practices (e.g Kelling and Coles, 1996) propose a zero-tolerance approach to minor
crimes, with particular emphasis on property crimes, loitering, public drunkenness,
and other misdemeanors. As Vitale’s (2008) work on New York’s “quality of life”
campaign suggests, the implementation of such programs can have considerable
impact on the daily operations of law enforcement. Though research has suggested
that the organization of law enforcement has a considerable effect on movement
repression (Cunningham, 2004), few studies have directly examined whether the
adoption of policies such as broken windows have influenced protest control.

While broken windows policing styles have been largely celebrated by law enforcement agencies and municipal governments, they have also generated a growing critical literature. Studies by Harcourt (2001) and Harcourt and Ludwig (2006), for example, suggest that the kernel of the broken windows position—that minimizing property crime and lower-level misdemeanors will reduce more serious crimes—is not empirically supported. Other work suggests that policies such as broken windows asymmetrically punish the poor, and particularly the homeless (Wacquant, 2009). More directly related to the issue of protest control is the implication broken windows policies can have on access to public space, and what law enforcement considers appropriate conduct in public spaces. McPhail et al.’s (1998) work on protest policing between between 1960 and 1995 points to an historical precedent where the police treated protest as a form of disorder, though they argue that over time the police came to recognize protest as a legitimate form of expression. Additionally, more contemporary evidence suggests that police control over public space, particularly during protest events, has increased dramatically (Noakes et al., 2005; Noakes and Gillham, 2006; Vitale, 2005, 2007; Zick, 2009). Such findings raise the possibility that the onset of programs such as broken windows might result in spillover effects, where protest is once again treated as a form of disorder. To date, this possibility has not been systematically examined by social movements scholars despite strong theoretical and empirical precedent.

2.1.2.2 Training and Tactical Strategies for Protest Control

Since the protest wave of the 1960s, there have been considerable changes in how the police are trained to handle large gatherings. As McAdam’s (1983) analysis of tactical innovation during the civil rights movement suggests, the interactions
between movements, counter-movements, and the state are quite dynamic. As a result, paying attention to prevailing trends in law enforcement—even if they do not appear to be directly related to protest policing—may help clarify or re-conceptualize the relationship between key causal variables.

Since the mid-1980’s, American police departments began to create Police Paramilitary Units (PPUs) to combat the war on drugs (Kraska and Paulsen, 1997). The remarkable growth of PPUs holds not only for large police departments (Kraska and Kappeler, 1997), but also for small and mid-sized departments as well (Kraska and Cubellis, 1997). Over time, PPUs came to be utilized by the police for proactive patrols, particularly when serving dangerous warrants (Kraska and Kappeler, 1997). One of the offshoots of the increasing reliance on PPUs for day-to-day police operations is that police officers receive training in Miami Field Force methods of crowd control, which emphasize acting collectively rather than independently or with a partner (McPhail and McCarthy, 2005). As a result, the police are able to much more efficiently handle large gatherings, even if they are vastly outnumbered.

The implications of the relatively new ability of the police to act in unison during demonstrations are vast since several studies have suggested that police presence and action is more likely at large events (e.g. Earl et al., 2003; Earl and Soule, 2006; Soule and Davenport, 2009). The mechanism behind this effect is that large events are more threatening to the police since larger gatherings have a higher capability to significantly disrupt public order. Since the bulk of the studies linking protest size to police action has looked at the period between 1960 and 1990, it is possible that this relationship is over-stated when it comes to contemporary protest. Research by Rafail (2010) on Canadian protest control between 1998 and 2004 has not replicated this finding. Furthermore, a recent study by Rafail et al.
(Forthcoming), which uses the same data source as the studies cited above, finds that the relationship between event size and police action is quite complex, and a simple positive relationship is not supported by the data. Because current training programs have enhanced the ability to contend with large gatherings, it is possible that well-attended demonstrations are no longer perceived as highly threatening.

Another series of studies have focused on how the police have increasingly tightened their spatial control over protest events by using old technologies such as barricades in innovative ways. Barricades and other similar physical structures are not new when it comes to protest control, however, they have increasingly been used in a repressive capacity by dividing the protestors into smaller, more manageable groups or to demarcate boundaries for free-speech zones (Zick, 2009). Such divide-and-conquer strategies have been used extensively by the New York Police Department (NYCLU, 2003; Vitale, 2005, 2007) and at many other major demonstrations (Gillham and Noakes, 2007; Noakes and Gillham, 2006). As Noakes et al. (2005) detail, barricades are now used to limit and control protestor mobility, monitor the protest site, and allow the police to preemptively or strategically force the protestors to alter their planned routes. While the costs required to construct and maintain complex barricaded structures are too prohibitive for universal adoption, they are relatively common at contemporary protest events. For instance, in the data used in this study, barricades were used at least 7% of the time overall, but at over 16% of protests in New York, NY.

2.1.2.3 Conservative Punitiveness

A final consideration revolves around the notion that punitiveness is highly contextual and to some extent, is related to the degree to which a particular region is politically conservative. For example, in a study that combines contextual and
individual correlates of sentencing, Helms and Jacobs (2003) demonstrate that conservative political contexts—as measured by the proportion of a county who voted Republican—lead to longer prison sentences. Additional work by Jacobs and Carmichael (2002, 2004) suggests that religious fundamentalism, political conservatism, and the strength of the Republican Party all significantly increase the prevalence of individuals sentenced to death. Overall then, more conservative regions are more likely to adopt punitive social control programs, even after statistically controlling for several other competing factors.

To date, few studies have examined whether protest policing is more aggressive in areas that are more conservative. This omission is curious given the regular use of violence and arrests by Southern police departments during the civil rights movement of the 1960s, coupled with the comparatively high level of support for Republican candidates in the South. Of the little research that has been done, the results suggest that local political structures do play a role in protest control. In their comparison of Zurich and Geneva, Wisler and Kriesi (1998) argue the way that political elites react to radical protest is differentiated by their understanding of the relationship between public order and constitutional order (see pp. 106–108). Little research has examined the U.S. case. As a result it is an open empirical question whether the linkage between conservatism and punitiveness can be applied to the case of protest policing.

2.1.3 Summary

To summarize briefly, I have argued that a considerable limitation in the scholarship on protest policing is that protest control is generally theorized to operated distinctly from state control more generally. As a solution, I have proposed fo-
focusing on variability in protest policing in cities rather than nations, which can
shed light on how structural factors influence protest policing. A conceptual di-
agram of the factors examined here is provided in Figure 2.1. While it is well
established that factors such as situational threats and protest-specific character-
istics influence protest policing, there has been little attention to the contextual
factors included in the figure. For example, the degree to which a police depart-
ment is experienced and trained to handle large crowds may indirectly influence
their response to demonstrations with many participants, which is indicated in
the figure. An undercurrent of several of proposed contextual effects is related to
increasing attention to the linkages between social stratification and punishment
(see e.g. Western, 2006), which suggests that factors such as income may also be
important predictors of aggressive policing, and potentially protest policing as well.
Programs including broken windows and other comparable crime control strate-
gies asymmetrically target those at bottom of the stratification system, suggesting
that the police may be more aggressive in communities that have comparatively
lower incomes. Given these considerations, I suggest that focusing on such meso-
structural factors creates a relatively concrete set of mechanisms to examine how
broader trends in social control systems affect protest policing, as well as the abil-
ity to compare and contrast protest policing practices with other trends in state
control. Such inter-relationships are quite difficult, if not impossible to examine
if one’s analytic focus is centered on movement-specific factors alone. I now turn
to a description of a unique database of contemporary protest events taking place
between 1996 and 2006 that allow for a direct assessment of how these factors all
contribute to the policing of protest, and the use of arrests in particular.
2.2 Data and Methods

To address my central research questions, I use a sample of protest events occurring in 20 major cities between 1996 and 2006, which is linked to data drawn from several sources of contextual information. The cities were selected from the 30 largest American municipalities, and the final list was selected to both maximize the geographic breadth of the sample while simultaneously ensuring that major centers of protest were retained.¹ A complete list of the cities I examine is provided in Figure 2.2.

The central source of data on protest events comes from newspaper coverage. The sources of contextual data for each city are described further below. The use

¹ This strategy has considerable advantages relative to selecting on city size alone. If size is the sole criteria for selection, it would result in a considerable over-representation of cities in Texas (over 20% of the sample), while central arteries of protest such as Washington DC would be omitted altogether.
of newspaper data has been criticized by some (e.g. Oliver and Myers, 1999; Oliver and Maney, 2000; Myers and Caniglia, 2004; Ortiz et al., 2005), largely because the mass media are much more (or less) likely to cover certain types of protest events—a phenomenon that McCarthy et al. (1996) refer to as selection bias. More specifically, several studies have suggested that newspapers and other media are more likely to devote attention to events that are large, particularly contentious, or otherwise remarkable, and this preference is consistent over time and across many different nation-states (McCarthy et al., 1996; Hocke, 1998; Barranco and Wisler, 1999; Earl et al., 2004; McCarthy et al., 2008). Additional research on coverage pattens of urban riots has found that that newspapers are much more likely to cover events occurring in close proximity to their editorial offices (Myers and Caniglia, 2004; Ortiz et al., 2005).

While a thorough discussion of selection bias and its implications is beyond the current scope, my sampling strategy provides a major corrective intended to limit the problems caused by selection bias. It is difficult to account for biases favoring relatively violent or contentious events, however my research design reduces selection bias resulting from geographic proximity: I use a major newspaper located in each of the cities under investigation (e.g. The New York Times in New York, the Boston Globe in Boston, etc.). \(^2\) This strategy departs from many other studies that have focused on a single newspaper such as the New York Times or the Washington Post, ultimately yielding a more rigorous sample. In the present case then, I side with Earl et al.’s (2004) review concluding that newspaper data provides an accessible way to collect protest event data that is acceptably valid and reliable.

\(^2\) When multiple newspapers were available for a particular city, I used the source with the highest annual circulation.
The focal unit of analysis is the protest event. To be included in the sample, each event must have been (A) publicly accessible; (B) taken place during the analytic period; (C) occurred in one of the cities under investigation; and (D) the participants must have aired a grievance. Articles on events were drawn from full-text searches with a broad search string, which returned over 1 million candidate articles. Full-text searches were used to avoid any inconsistencies in topical indices across different newspapers. After following these conventions, the final analytic sample contains 11,426 protest events that were coded from more than 13,753 articles.

2.2.1 Variables and Operationalization

The focal response variable measures whether any arrests took place. Similar to many other empirical studies of protest policing, this variable is operationalized as dichotomous, capturing whether (= 1) or not (= 0) any arrests took place. Clearly there are multiple ways to operationalize protest policing, yet there are several distinct advantages to focusing on arrests in this analysis. First, being arrested is merely the initial step into the criminal justice system, and this process can last many years for activists even if the charges are eventually dropped (Earl, 2005). Excluding cases of lasting bodily harm or psychological trauma, this is generally not the case when the police use other forms of physical force. Second, unlike the protest wave of the 1960s where the police relied on violence as their primary means to control demonstrations, there has been a historical shift away from such strategies (McPhail et al., 1998). Despite this, arrests have remained a staple of protest policing strategies over time (Rafail et al., Forthcoming). Third, the behavior of police at recent high profile protests, ranging from the 1999 WTO
protest in Seattle to the 2004 Republican National Convention protests in New York, suggests that the police may rely on arrests as a demobilization tool. A report by the New York Civil Liberties Union (NYCLU, 2003) focusing on the conduct of the New York Police Department during a major anti-war demonstration in 2003 reached largely similar conclusions. Such heavy-handed use of arrests may signal to potential participants that even legally demonstrating carries the risk of arrest. For these reasons, I suggest that focusing on arrests provides a feasible, though incomplete measure of protest policing.

A total of four substantive blocks of predictor variables are used in the statistical analysis. These include situational threats, protest characteristics, other police tactics, and several contextual, city-level indicators. I now turn to their operationalization.

I use four measures of situational threats: First, civil disobedience is a measure for whether the protestors non-violently disrupt public order or violate the law. Examples include sit-ins, camp-outs, or blocking traffic. Second, I use an indicator for a contest over space, which occurs when the participants attempt to or succeed in breaching an area that is out-of-bounds (e.g. protesting inside a business), or otherwise using space to challenge law enforcement control over a protest site. This variable differs from civil disobedience in that it has an explicitly spatial component. Third, property damage refers to cases where the demonstrators break windows, destroy merchandise, or otherwise damage public or private property. Finally, the use of violence measures any event where the protestors threaten or engage in violent activities including assaulting the police or other bystanders. All of these variables are operationalized as dichotomous and are coded as present ( = 1) or absent ( = 0).

Three measures are used for protest characteristics. These are non-behavioral
elements of a protest event that are legal but may nonetheless influence police behavior. First, I use a dummy variable indicating whether a counter-movement is present (= 1) or not (= 0). Second, I assess whether a particular social movement organization (SMO) is reported to be present. This is because previous studies have suggested that SMOs are more likely to form cooperative relationships with the police (McPhail et al., 1998). Finally, I use the natural log of the number of participants reported to be present at an event.\(^3\) When multiple size estimates were given, for example from the police and the protestors, the mean of all of the estimates is used.

Two measures of police tactics are used as predictor variables. First, I use an indicator for whether the police used any sort of barricade during the protest. These can include steel, wooden, or other physical structures as well as lines of police officers intended to limit the mobility of the demonstrators. Second, the use of force measures when the police use any sort of physical force, less-than-lethal weapons, or other equipment during the protest event.\(^4\) Both variables are coded as dichotomous indicators (0 or 1).

To expand the scope of this analysis beyond protest-level dynamics, I use a series of contextual factors theoretically linked to protest policing. These variables are generally measured annually and are drawn from official databases. I use these measures as level-2 covariates in the regression analysis described further.

---

\(^{3}\) Following Rafail et al. (Forthcoming), when a precise numerical estimate is not provided I use a uniform probability distribution with cut-points representing the beginning and ending of a categorical estimate of the crowd size that is based on textual clues provided in the article. These values are used to patch in the missing data. Unlike mean imputation, this retains the mean of the observed distribution, but does not artificially decrease the variance.

\(^{4}\) Arrests are also a use of police force, however, I suggest that it is useful to keep arrests and other forms of force distinct. Since arrests mark the starting point of what my be a prolonged interaction with the criminal justice system (Earl, 2005), the actions of the police while making arrests (e.g. Miranda rights, due process) are likely to scrutinized. This is not necessarily the case for other forms of force. As a result, the police may be more cautious when deciding to make arrests.
below. The first covariate measures the annual *median income* for the metropolitan statistical area of each city. This data is provided by the Bureau of Economic Analysis, and is available on the World Wide Web (http://www.bea.gov). Due to the magnitude of this variable, it is scaled to units of 10,000.

Annual measures of crime rates in each city are drawn from the Uniform Crime Reports (UCR) series that is collected by the Federal Bureau of Investigation. These are reported by the municipal police department for each city in the sample. The FBI differentiates between *violent crime* and *property crime*, and these distinctions are also used here. The crime rates are scaled so that they reflect the number of crimes that are cleared per 100,000 civilians. Since broken windows philosophies emphasize aggressive policing of property crimes, I use these variables as a proxy measure of the degree to which a particular police department has implemented such policies. While another strategy to operationalize this concept would be to use a dichotomous variable indicating whether a particular police department has adopted such policies, previous literature suggests that police discourses need not reflect their organizational practices (Kappeler and Kraska, 1998; Zhao et al., 2001). I use the number of violent and property crimes cleared (i.e. solved) per 100,000 civilians rather than the number of crimes *known* to the police, since clearance rates may better reflect departmental priorities. Despite this, the results are substantively similar when the crime rates known to law enforcement are used making this choice of little consequence. Finally, since previous research by Earl and Soule (2006) has suggested that police capacity is linked to protest policing patterns, I also include a measure of the number of sworn police officers for each department in the sample. This data is drawn from the UCR series and is scaled so that it reflects the number of sworn police officers per 100,000 civilians.

To assess the political climate of each city and its surrounding region, I use
data on the House of Representative electoral contests taking place in 1994, 1996, 1998, 2000, 2002, and 2004. Figures for the House of Representatives elections were chosen because they occur more regularly than Senatorial or Presidential contests and there is a higher overall number of House candidates. Federal level elections are used because the House electoral processes are comparable across the US, unlike the wide range of municipal election systems that are not easily standardized. The official tallies of the House elections were compiled from the CQ Press Elections and Voting Collection. I use a variable for the proportion of the vote in favor of the Democratic Party, which is lagged one election cycle in all models.

2.2.2 Analytic Strategy

Bayesian semiparametric multilevel logistic regression models are used to examine patterns of arrests. There are two advantages to working from the Bayesian paradigm: First, frequentist approaches to generalized multilevel models typically rely on either Penalized Quasi-Likelihood (PQL) or LaPlacian approximations of the likelihood function (Raudenbush and Bryk, 2002). Several studies have suggested that such procedures may underestimate the magnitude of variance parameters for random effects or have higher than expected levels of mean-squared error (see e.g. Bellamy et al., 2005; Diaz, 2007; McCulloch and Searle, 2001). In contrast, such concerns are avoided altogether using Bayesian estimation via Markov chain Monte Carlo (MCMC) techniques (Neuhaus and Segal, 1997; Browne and Draper, 2006). Since examining the city-level variance is a key goal of this research, it is likely that Bayesian methods provide either equivalent or better results relative to the frequentist estimators. Second, as Rafail (2010) notes, Bayesian methodology can provide a much more intuitive inferential engine when dealing with newspaper
data since the results can be interpreted directly. In contrast, more commonly used $p$-value based statistical inference requires the assumption that the sampling mechanism can be replicated (Jackman, 2009; Gill, 2008), which is rarely the case for newspaper data.

Since previous research on protest policing by Rafail et al. (Forthcoming) has suggested that assuming that continuous predictor variables—particularly event size in the present case—have linear monotonic effects may be inappropriate, I use semiparametric analysis. Such procedures allow a much more flexible, though computationally intensive, approach to dealing with non-linear relationships in a regression model (see Fox, 2000; Keele, 2008; Andersen, 2009). An iterative process was used to identify potential non-linear relationships, where initially all of the continuous variables were specified as nonparametric components of the model. The results were visually examined and ultimately, only the effect of event size had a clearly non-linear effect on arrests, and a linear fit was appropriate for the other continuous variables. The results below are based on a model where only the variable for event size is entered as a nonparametric term and the other continuous variables are treated as standard parametric effects.

Conceptually, I treat protest events (level 1) as nested in city-years (level 2), and therefore specify two-level regression models. This strategy allows the introduction of covariates that vary not only by city but also over the analytic period allowing a more dynamic interpretation of the structural covariates. I use a varying-intercept model, which allows for a random effect parameter for each city-year. This approach allows for the simultaneous estimation of protest level

---

5 For a small subset of the cities I examine (Detroit, Indianapolis, Nashville, and Phoenix), full-text newspaper coverage was unavailable between 1996 and 1998. To correct this examined newspaper coverage of protest events for all available newspaper in the state where each city was located, however, the coverage patterns pointed to significantly fewer covered events. As a result, I only analyze 1999 through 2006 for these cities.
factors (i.e. situational threats, protest characteristics, and police tactics) as well as the contextual measures described above. Since the models are Bayesian, prior information must be included. Each regression coefficient is assigned a noninformative prior distribution such that \( \Pr(\hat{\beta}) \sim \mathcal{N}(0, 0.001) \). Following Ruppert et al. (2003) and Crainiceanu et al. (2005), I use evenly spaced quantile-based knot points for nonparametric smoothing. A total of twenty knots are used. These knots are parameterized as random effects in the multilevel model, and are also assigned normal prior distributions with a mean of 0 and precision of 0.001. I use 100,000 Markov chain Monte Carlo iterations with two parallel chains to simulate the joint posterior distribution of the regression parameters. The first 10% of the iterations were treated as burn-in values and were discarded. A variety of diagnostics were used to assess chain convergence including Gelman-Rubin statistics, Geweke tests, autocorrelation plots, and trace plots (see Jackman, 2009). These measures all pointed towards satisfactory convergence. In equation format, the full regression model is

\[
\text{Logit} (\text{Arrests}) = X\beta + X\beta + f(\text{Size}) + X\beta + X\beta + Zb .
\]

\( X \) is used to denote matrices of predictor variables, \( \beta \) for regression coefficients, \( f(\cdot) \) for a nonparametric smooth, \( Z \) for an indicator matrix for each city-year, and \( b \) for a vector of random effects.

\( \text{6 Consistent with the Bayesian literature, I use precisions, } \tau^2, \text{ rather than variances, } \sigma^2. \text{ The variance is simply } \tau^{-2} = \sigma^2. \text{ This is equivalent to specifying a normal prior distribution with a mean of zero and standard deviation of 1,000. On the logit scale, this clearly denotes highly noninformatiave priors.} \)
Figure 2.2. The occurrence of arrests at protest events taking place in 20 major U.S. cities, 1996-2006 ($n = 11,426$)

2.3 Results

I begin by summarizing the baseline differences in arrests and descriptive statistics for the predictor variables. Figure 2.2 provides the percentage of events where arrests took place for each of the cities examined. For the purposes of comparison, the average percentage of arrests for all cities is provided as well. Table 2.1 provides the descriptive statistics.

It is clear in Figure 2.2 that there is tremendous diversity across the twenty cities. Indeed, in New York, NY, over 20% of events have arrests, while less arrests take place at fewer than 5% of protests in Nashville or Louisville. Comparing the
Table 2.1. Descriptive statistics for predictor variables

<table>
<thead>
<tr>
<th>Event Level Predictors</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Civil Disobedience</td>
<td>0.10</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Contest Over Space</td>
<td>0.09</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Property Damage</td>
<td>0.08</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Use of Violence</td>
<td>0.05</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Counter-Demonstrators Present</td>
<td>0.12</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>SMO Present</td>
<td>0.55</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Size of Event (ln)</td>
<td>4.45</td>
<td>1.97</td>
<td>0.00</td>
<td>13.82</td>
</tr>
<tr>
<td>Police Use of Barricades</td>
<td>0.07</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Police Use of Force</td>
<td>0.04</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Contextual Predictors</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median Income ($10,000)</td>
<td>3.43</td>
<td>0.58</td>
<td>2.25</td>
<td>5.36</td>
</tr>
<tr>
<td>Property Crime Rate (100,000)</td>
<td>6529.30</td>
<td>2103.08</td>
<td>1009.02</td>
<td>15129.90</td>
</tr>
<tr>
<td>Violent Crime Rate (100,000)</td>
<td>2427.80</td>
<td>1266.31</td>
<td>242.24</td>
<td>7012.63</td>
</tr>
<tr>
<td>Proportion Voted Democrat</td>
<td>0.60</td>
<td>0.17</td>
<td>0.25</td>
<td>0.99</td>
</tr>
<tr>
<td>Police Capacity (10,000)</td>
<td>28.09</td>
<td>14.13</td>
<td>1.70</td>
<td>56.00</td>
</tr>
</tbody>
</table>

Note: n’s are 11,234 for protest events and 205 for contextual effects.

spread of the distribution relative to national average from all the cities strongly suggests that focusing on more aggregate units of analysis alone masks considerable diversity in protest policing within a single nation-state. Such descriptive results are limited, however, since they do not account for variability in the types of protest events which the police must contend with. For instance, if protest events in New York or San Francisco are particularly contentious then we should expect higher levels of arrests, other things being equal.

The descriptive statistics in Table 2.1 point to several interesting trends. Of particular interest are the averages of several key measures. For instance, the more contentious protestor tactics occur at 10% or less of the events, and protestor initiated violence in particular is relatively infrequent. This is consistent with claims by McCarthy and McPhail (1998) and others that social movement mobilization has become less contentious as protest has become institutionalized in the United States. The police use of force is also rare, occurring at 4% of events. Comparing
the results in Figure 2.2 and Table 2.1 it is clear that arrests are, on average, much more prevalent than other uses of force in contemporary protest policing. To assess such relationships more systematically, I now turn to the results from the regression analysis, which allows for a direct test of whether such pronounced city-level variation remains once the variables for situational threats, protest characteristics, other police tactics, and contextual factors are included.

The regression analysis is based on a comparison of three model specifications: first, I use a single-level logistic regression model containing only the protest-level covariates. This approach replicates the most common strategy in the existing literature because it ignores any potential impacts of contextual factors or city-level variability. Second, to account for city-level variability, I introduce a random effects parameter for the intercept using a multilevel model. Third, I add a matrix of contextual predictor variables for each city-year. The Deviance Information Criterion (DIC) for each model is respectively 3256.50, 3168.53, and 3124.07.\(^7\) Based on this it is clear that the full model considerably improves fit. I focus on the posterior distribution of the full regression model, though a comparison of these models is appended as Table 2.2.

An advantage of the Bayesian paradigm is that the analyst can more naturally focus on distributions of coefficients rather than simple point estimates of coefficients and standard errors (Gelman et al., 2004; Lynch, 2007; Gill, 2008). In what follows, I summarize the results from the posterior distributions graphically. The posterior means in Figure 2.3 through Figure 2.6 are analogous to a coefficient estimate from frequentist analyses, while 95% credible intervals are roughly anal-

---

\(^7\) The DIC is a Bayesian measure of fit proposed by Spiegelhalter et al. (2002) which is a numerical summary of model fit after taking complexity into account. Lower values of the DIC suggest better fitting models, and a change in DIC of 10 or more indicates a considerable improvement in the explanatory power of a model.
Figure 2.3. Posterior means and 95% credible intervals for situational threats from a Bayesian semiparametric multilevel logistic regression model predicting arrests during protests in 20 U.S. cities, 1996-2006 ($n = 11,426$)

ogous to the more widely used 95% confidence intervals. Unlike their frequentist counterparts, however, the Bayesian variants have a much more intuitive interpretation: given the data, we can be 95% confident that a particular effect is positive, negative, or both. This is not the case for frequentist inference, where probabilities must be interpreted with reference to an theoretical infinite number of trials (Jackman, 2009). To summarize the results, I present each block of predictor variables individually, though the results are drawn from the full model in Table 2.2.

Figure 2.3 summarizes the posterior distributions for situational threats. As expected, the posterior means and credible intervals are uniformly positive, sug-
gesting that when faced with contentious and illegal conduct, the police are much more likely to make arrests. Overall, these effects further confirm that situational threats are of considerable importance in understanding protest policing. It is important to note that these estimates remain strong and positive even after including the other predictor variables as well as a varying-intercept parameter for each city-year. These results further corroborate the finding in several previous studies by Earl et al. (2003), Soule and Davenport (2009), and others that situational threats are strong and robust predictors of police action during demonstrations. Despite this, turning now to the results of protest characteristics, it is also clear that situational threats alone do not provide an exhaustive explanation for protest policing.

Figure 2.4 contains the estimates for protest characteristics. Panel A contains the 95% credible intervals for the presence of counter-demonstrators and SMO presence. Panel B provides the nonparametric smooth is represented by a solid line for the posterior mean and dashed lines for the lower and upper boundaries of the 95% credible interval. Based on Figure 2.4A, the presence of counter-demonstrators decreases the likelihood of arrests. This suggests that when two or more competing or antagonistic groups are in close proximity the police are more prone to follow a containment strategy rather than more aggressive crowd control tactics. Compared to the findings reported by Earl and Soule (2006) and Soule and Davenport (2009), which had a positive association between counter-demonstrators and police action, this analysis supports the hypothesis that the police have indeed changed their strategies. The posterior mean for SMO presence is nearly zero, and its credible interval is between roughly −0.20 and 0.20. While previous literature (e.g. McPhail et al., 1998) has suggested that the police may develop cooperative relationships with SMOs, resulting in a net advantage to organizational presence, this does
Figure 2.4. Posterior means and 95% credible intervals for protest characteristics from a Bayesian semiparametric multilevel logistic regression model predicting arrests during protests in 20 U.S. cities, 1996-2006 ($n = 11,426$)

A: Parametric Effects

B: Nonparametric Smooth

not appear to be supported in these data. A potential explanation is that SMOs have been increasingly common as social movements have become institutionalized (McCarthy and McPhail, 1998), and as a result, the commonality of organizational sponsorship outweighs any advantages.

The nonparametric smooth in Figure 2.4B indicates that there is a complex relationship between event size and the use of arrests. We see a decrease in the logit of arrests as event size grows, however this trajectory declines and flattens at approximately a log of 4. The estimate then slowly, but monotonically increases until a log of around 11, then begins to decline again. The credible interval also widens as the event size increases. For the largest events, the police have no clearly preferred strategy, and the 95% credible interval is at its widest point for these events. Overall, these result confirm that even if the police treated large demonstrations as threatening in the past, this no longer seems to be the case.

The estimates for police tactics are provided in Figure 2.5. The posterior means and credible intervals for both the use of barricades and the use of force are positive,
Figure 2.5. Posterior means and 95% credible intervals for police tactics from a Bayesian semiparametric multilevel logistic regression model predicting arrests during protests in 20 U.S. cities, 1996-2006 ($n = 11,426$)

though the magnitude of effect size for the latter is much larger. Nonetheless, this supports the previous studies by Noakes et al. (2005), Vitale (2005, 2007), Zick (2009) and others that barricades are used in a repressive capacity, not simply as a tactic used to promote the safety of the participants by constraining their mobility. That there is an increase in arrests when the police use force is also substantively significant since McPhail et al.’s (1998) work on Washington DC hypothesizes a decoupling of violence from arrests.

Figure 2.6 contains both the predicted probabilities of arrests in Panel A and the posterior densities for the contextual effects in Panel B. The results in Panel A are
Figure 2.6. Posterior means and 95% credible intervals for contextual factors from a Bayesian semiparametric multilevel logistic regression model predicting arrests during protests in 20 U.S. cities, 1996-2006 (\(n = 11,426\))

**A:** Observed (red) and predicted (black) city-level variation

**B:** Contextual effects

constructed by creating city-specific averages based on the posterior distributions for each random effect estimate (i.e., each city-year). Relative to the observed sample proportions (the red dots in Figure 2.6A), the model predicts the underlying differences between the cities very well (see the black dots and lines). A minor exception to this close correspondence is for Detroit, where the model overestimates the probability of arrests by approximately 4%. This is likely because there were only 211 events coded for Detroit over the analytic period, ultimately providing less information required for inference. Despite this, the observed proportion for Detroit is still well within the expected range. More generally, the pronounced city-level differences demonstrate that even after controlling for situational threats, other protest characteristics, and other police tactics, there remains considerable diversity across the United States.

The results for the contextual effects are provided in Figure 2.6B. There is a negative relationship between median income and arrests, indicating that the
police privilege events taking place in wealthier locations. This is consistent with the research linking higher levels of state control directed at those residing in less wealthy communities, and suggests that such dynamics may also influence protest policing. The estimates for the effect of violent and property crime clearance rates are also noteworthy. While there is an increase in the likelihood of arrests when the police have higher property crime clearance rate, we cannot reach any definitive conclusions about the impact of violent crime clearance rates. Since programs such as broken windows are predicated on law enforcement agencies allocating resources towards limiting property crimes, this suggests that such police philosophies do indeed spill over into protest control. Similar to Oliver’s (2008) analysis, these results also support the idea that protest repression is linked to crime control, and perhaps, that broken windows policies increases police aggressivity during protest events. Taking the effects for income and property crime together, there is also preliminary evidence that not only do policies such as broken windows target the poor, but they may also curtail First Amendment expression in less wealthy areas. The credible interval for the proportion voted democrat is also uniformly positive, though of a relatively small magnitude. This effect is in the opposite direction that was expected, and suggests that the linkage between political conservatism and punitiveness is fundamentally different for protest events. A more thorough discussion of this finding and its implications is provided in the concluding section. Finally, the estimate for police capacity has a positive posterior mean, however, the 95% credible interval overlaps zero making it difficult to decisively conclude that police capacity increases or decreases the likelihood of arrests. Even though one might expect that police departments with more labor power might also have more resources to allocate to control protest events, this does not appear to be supported when looking at arrests. Instead, it appears that high levels of police
capacity does not lead to more aggressive protest control if the willingness to make arrests is not present.

2.4 Discussion and Conclusion

This study raises several implications for future research. First, the main implication of the results supports the argument that movement control has been conceptualized too narrowly, and that its evidentiary scope has likely been constructed too broadly. It is clear that there is considerable variability in the way that the police handle protest events across the United States, and that different police departments have significantly different public order management strategies. It follows that research focusing on more aggregated units of analysis such as states or nations, over-state the uniformity of protest policing, both theoretically and empirically. Though most of the variables measuring situational threats and protest characteristics were in similar directions as previous work (e.g. Earl et al., 2003; Soule and Davenport, 2009), the inclusion of departmental availability and contextual factors extends the analytic scope of protest policing theory in a way that allows explicit linkages to other research on policing, punishment, and state sponsored social control more generally. That is, the observed variability in protest policing can be partially explained by examining the municipal contexts where protest occurs. This has largely escaped the attention of social movements scholars. These results suggest not only that patterns of crime control and other contextual factors are important to a theory of protest policing, but focusing on the local level also enhances our capacity to evaluate their impact. Future research should pay attention to these dynamics, particularly since the traditional focus on national-level protest control completely overlooks such influences.
While many of the protest-specific variables were comparable to previous research, a second major implication relates to the relationship between event size and protest policing. While previous literature has typically found that the police are more aggressive when faced with larger events, the nonparametric smooth reported in Figure 2.4B points to a much more complex relationship. Other work by Rafail et al. (Forthcoming), which also uses semiparametric modeling, reaches similar conclusions regarding the problems of conceptualizing these relationships as simple and monotonic (though the overall shape of the smooths are different). Nevertheless, such results minimally raise the issue of whether the widely found effects of event size are due to overly strict and incorrect parametric assumptions used. The effect of event size, particularly when coupled with the negative relationship between counter-demonstrator presence and arrests lends support to the idea that current police officer training may reduce the level of perceived threat attributed to demonstrations that have traditionally been policed more aggressively.

Third, this research demonstrates that there is considerable variability in protest control, however, the number of cities I examined makes it difficult to more concretely engage with the causal underpinnings of this variation. While it is beyond the scope of this paper to provide a comprehensive explanation of the factors influencing protest policing, a more thorough examination with a smaller number of cities could provide the beginnings of a theoretical model attempting to understand how protest policing practices in a given municipality are constructed and maintained. Ethnographic analyses, similar to Waddington’s (1994) work on the Metropolitan Police Service in London, could be an ideal venue to develop a deeper understanding of how local law enforcement, political, and economic contexts influence protest control. Nonetheless, establishing that locally grounded but structured variability exists in protest policing is an important first step, and
future research should more systematically pursue its sources.

A fourth point raised by this research is that I focus solely on overt forms of protest control initiated by state actors. This is an incomplete picture of the totality of the tactics used by agents of the state to repress social movements. Future research should address the covert forms of protest policing with particular emphasis on whether their have been any change in information gathering strategies since the 9/11 attacks on the United States. As della Porta and Reiter (1998) discuss, intelligence gathering is a major component of protest policing, and it is important to evaluate whether variation exists in police practices, and if so, how it is tied to structural conditions. In addition, a comparison of the structural and contextual features that foster the growth of right wing social movements may also shed light on private repression of social movements.

Finally, the higher likelihood of arrests in more liberal cities also requires commentary since this relationship is in the opposite direction of what has been found in previous research on social control (Helms and Jacobs, 2003; Jacobs and Carmichael, 2002, 2004). This finding is quite significant, particularly for scholars interested in how political opportunity structures influence social movement mobilization. The Democratic Party has traditionally been treated as the major political party that provides increased opportunities to many progressive movements (e.g. the pro-choice or gay rights movements), but it appears as though such opportunities do not carry over to more tolerant protest policing. One explanation for this effect is that more liberal cities also have more aggressive protest cultures, which may also result in more proactive policing. Supplementary analysis (available upon request) confirms that the cities with higher levels of democratic support also have more events with civil disobedience, contests over space, property damage, and violence. Cities in the analytic sample that are more liberal also
have a higher frequency of protest events and are more diffuse in the claims made by the challenging groups. Overall then, there are more repressive opportunities and contentious histories in cities with higher support for the Democrats and this may partially explain the counter-intuitive effects.
Table 2.2. Bayesian logistic and multilevel logistic regression posterior means and standard deviations predicting arrests at U.S. protest events, 1996–2006

<table>
<thead>
<tr>
<th></th>
<th>Event Characteristics</th>
<th>Varying Intercept</th>
<th>Contextual Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Protestor Tactics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Civil Disobedience</td>
<td>2.29</td>
<td>2.42</td>
<td>2.42</td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
<td>(0.13)</td>
<td>(0.13)</td>
</tr>
<tr>
<td>Contest Over Space</td>
<td>1.39</td>
<td>1.37</td>
<td>1.31</td>
</tr>
<tr>
<td></td>
<td>(0.13)</td>
<td>(0.14)</td>
<td>(0.14)</td>
</tr>
<tr>
<td>Property Damage</td>
<td>1.07</td>
<td>1.09</td>
<td>1.09</td>
</tr>
<tr>
<td></td>
<td>(0.21)</td>
<td>(0.22)</td>
<td>(0.22)</td>
</tr>
<tr>
<td>Use of Violence</td>
<td>2.35</td>
<td>2.32</td>
<td>2.29</td>
</tr>
<tr>
<td></td>
<td>(0.17)</td>
<td>(0.17)</td>
<td>(0.18)</td>
</tr>
<tr>
<td><strong>Protest Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Counter-Demonstrators Present</td>
<td>-0.56</td>
<td>-0.49</td>
<td>-0.45</td>
</tr>
<tr>
<td></td>
<td>(0.15)</td>
<td>(0.15)</td>
<td>(0.15)</td>
</tr>
<tr>
<td>SMO Present</td>
<td>-0.03</td>
<td>-0.02</td>
<td>3.4E-4</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.10)</td>
<td>(0.10)</td>
</tr>
<tr>
<td>Size of Event (ln)</td>
<td>Fig. 2.4B</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Police Tactics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of Barricades</td>
<td>0.62</td>
<td>0.53</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>(0.17)</td>
<td>(0.18)</td>
<td>(0.18)</td>
</tr>
<tr>
<td>Use of Force</td>
<td>3.10</td>
<td>3.27</td>
<td>3.28</td>
</tr>
<tr>
<td></td>
<td>(0.19)</td>
<td>(0.21)</td>
<td>(0.20)</td>
</tr>
<tr>
<td><strong>Contextual Effects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median Income ($10,000)</td>
<td>-0.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Property Crime Rate (100,000)</td>
<td>0.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.32)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Violent Crime Rate (100,000)</td>
<td>-0.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.71)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion Voted Democrat</td>
<td>0.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Police Capacity (100,000)</td>
<td>0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-3.23</td>
<td>-3.48</td>
<td>-3.79</td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(0.15)</td>
<td>(0.43)</td>
</tr>
<tr>
<td>Variance Component, $\sigma^2_\alpha$</td>
<td>0.38</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.09)</td>
<td></td>
</tr>
<tr>
<td>Deviance Information Criterion</td>
<td>3256.50</td>
<td>3168.53</td>
<td>3124.07</td>
</tr>
</tbody>
</table>
Chapter 3

Policy Spillover and the Policing of Protest in New York City, 1960-2006

3.1 Introduction

Analyses of long-term trends in the policing of protest has been a topic of considerable interest to social movements scholars. Within this literature several studies have focused on how police approaches to protest management have changed since the wave of activism in the 1960s (e.g. Soule and Davenport, 2009; Davenport et al., 2011; Rafail et al., Forthcoming). Sustained scholarly interest in this topic is not surprising since the relative severity or permissiveness of the dominant protest policing strategies in a society can have considerable implications for contentious political action as a whole. In her widely cited work on Germany and Italy for example, della Porta (1995) argues that protest policing is a barometer of a society’s broader political opportunity structure, implying that key areas affecting social movement mobilization (e.g. access to elite actors) are somewhat contingent on protest policing.
Among the most widely accepted explanations for trends in U.S. protest policing is the account provided by McPhail et al. (1998), though other accountings also exist (see Goldstein, 2001; Soule and Davenport, 2009). McPhail and colleagues differentiate between two major periods of protest policing—*escalated force* and *negotiated management*—that collectively represent a shift from protest control by force towards a negotiation-based model of interaction intended to minimize conflict during collective action events. McPhail et al. (1998) based on their arguments on observing dozens of encounters between the police and protestors in Washington D.C., cautioning that their findings may not generalize evenly to other locations. Few studies have examined the possible reasons how or why the evolution of protest policing in a particular location might deviate from the patterns in Washington D.C., providing a significant gap in scholarly work during this important period of transition.

This study provides an alternative explanation of how protest policing evolved in New York, NY, drawing on a sample of 6,147 protest events occurring between 1960 and 2006. This accounting is based on two observations pointing to limitations of the existing theoretical explanations: First, protest policing encompasses only a small portion of police activities, yet the dominant explanations of protest policing rely mainly on social movement dynamics to explain law enforcement actions. A viable accounting of protest policing must also attend to the political construction of crime and criminality and how these issues affect the social organization of law enforcement. Second, there is a growing tendency in the literature on protest policing emphasizing the impact of threat, referring mainly to the behavior of protestors during an event. While the importance of threat cannot be denied (see Earl et al., 2003; Earl and Soule, 2006; Soule and Davenport, 2009), emphasizing the behavior of the protestors above all else decontextualizes how the police may
come to privilege one group over another, which I have elsewhere termed “group-specific threats” (Rafail, 2010). This explanation hinges on what I refer to as *policy spillover*, or how policy decisions not directly related to protest policing nonetheless result in observable differences in how the police handle demonstrations. New York is an excellent location for a case study since the city adopted several major policy and institutional changes between 1960 and 2006, some of which has been linked to protest policing (Vitale, 2005, 2007). I particularly draw on crime and crime control in New York as central explanatory factors influencing protest policing there over time.

To proceed I first outline McPhail et al.’s (1998) historical argument of how existing protest policing practices have evolved from a basis in coercion and aggression to a relatively more accommodating and inclusive model. I then critique this explanation by drawing on the apparent contradiction between the expected softening of protest policing and the New York Police Department’s parallel campaign of zero-tolerance policing. Drawing on a series of changes implemented in the late 1970s and early 1980s by the administration of Mayor Edward Koch that had a lasting legacy on New York City politics and institutions. I focus specifically on the effects of spatial privatization and the adoption of Broken Windows policing and argue that together these factors had latent effects on protest control and ultimately launched a much less tolerant form of protest policing. Next, I describe my data and statistical methodology and then turn to my results. I conclude by discussing the implications of this research as well as how policy spillover may influence protest policing in places other than New York.
3.2 The Control of Protest Since the 1960s

The wave of social movement mobilization starting in the 1960s was marked by highly coercive state responses, ranging from the egregious show of force during the Democratic National Convention in 1968 held in Chicago (Walker, 1968) to the systematic acts of brutality directed at civil rights activists (Morris, 1986). Since then many scholars have suggested there has been a softening of police tactics and more permissive approaches to protest management were increasingly axiomatic (e.g. McCarthy and McPhail, 1998; della Porta and Fillieule, 2004; Soule and Earl, 2005). This argument is not without its critics, with several examples of highly coercive police responses to protest such as 2004’s Republican National Convention meetings in New York (Earl, 2009), the policing of major trade meetings (Gillham and Noakes, 2007; Noakes and Gillham, 2006; Noakes et al., 2005), and the targeting of certain types of protestors over others regardless of their behavior (Rafail, 2010; Davenport et al., 2011). There is a general consensus, however, that at least for the period between 1970 and the late 1980s, the police were more hesitant to use overt forms of violence during demonstrations. To examine the underlying mechanisms resulting in the movement away from coercion-based protest policing I now turn to McPhail et al.’s (1998) argument.

3.2.1 From Escalated Force to Negotiated Management

McPhail et al.’s (1998) provide a synthetic, historical argument distinguishing between two major styles of protest policing. The first, escalated force, is grounded in coercion and force as the primary protest management strategies and was the standard police response between 1960 and the late 1970s. The second style is referred to as the negotiated management model, which is based on mutual com-
munication and cooperation between the protestors and the police. They suggest
that negotiated management became the dominant approach to protest control in
the early 1980s and it remains the standard model of police conduct in the present.
There are five major dimensions that differentiate escalated force from negotiated
management, which are treated as continua rather than absolute contrasts. The
key dimensions and differences between the two models of protest control are sum-
marized in Table 3.1, which is adapted from McPhail et al. (1998). It is clear
from the table that negotiated management is more permissive. Indeed, relative
to the escalated force period, the police are more likely to recognize protest as a
legitimate and protected form of political expression.

Table 3.1. Key dimensions and differences between the Escalated Force and Negotiated
Management models of protest policing

<table>
<thead>
<tr>
<th></th>
<th>Escalated Force</th>
<th>Negotiated Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concern for 1st Amendment Rights</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Tolerance for Community Disruption</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Communication between Police and Protestors</td>
<td>Rare</td>
<td>Frequent</td>
</tr>
<tr>
<td>Extent and Manner of Arrests</td>
<td>Frequent</td>
<td>Rare</td>
</tr>
<tr>
<td>Extent and Manner of Force</td>
<td>Frequent</td>
<td>Rare</td>
</tr>
</tbody>
</table>

To explain the transition from escalated force to negotiated management, McPhail
et al. (1998) develop the concept of Public Order Management Systems (POMS),
referring to the organizational structures, legal environments, and technologies
that resulted in the development of protest policing policies. They identify four
mechanisms that collectively resulted in abandonment of escalated force and the
ascendancy of negotiated management. First, the wave of riots and civil distur-
bances of the 1960s and 1970s (see Myers, 1997, 2000; Perez et al., 2003; Myers,
2010, for details) resulted in a several National Commissions designed to under-
stand and minimize future uprisings. The recommendations of the commissions
included improving civil disorder training, using the minimal amount of force necessary during events, and emphasizing that protest is a legitimate form of political expression that ought to be protected. Second, as a consequence of the regular use of arrests and force during the escalated force period a series of legal challenges made their way to the Supreme Court. This resulted in the codification of a right to assemble and provided explicit guidelines, at least in principle, for when restrictions on First Amendment activity were permissible under the law (though for discussion on more recent developments see McCarthy and McPhail, 2006; Zick, 2009). Third, there was a widespread development of protest permitting systems across the United States. Research on permitting systems remains rare (but see Martinez, 2011) though there is little question that they significantly reduced tensions between protestors and the police and opened channels of communication between them. Finally, the United States Military Police School developed training courses in civil disorder policing that provided a series of tactics and strategies that were widely adopted by local law enforcement. The courses increased municipal, state, and county police department’s ability to prepare for and control riots and demonstrations and emphasized the show of force rather than the use of force.

The contrast between the escalated force and negotiated management models is explicitly ideal typical and McPhail et al. (1998) point out that their categorization is relative and has several exceptions, notably arising from uneven diffusion of negotiated management practices to municipal police departments across the United States. They also note that police discretion remains a core component of protest control, resulting in many examples of police responses that may seem out of character given standard operating procedures. For example, in Earl et al.’s (2003) examination of the policing of protest in New York State between 1968 and 1973—the height of the escalated force period—the police were not reported to be
present at the majority of events.

3.2.2 The Case for Divergence: Protest Policing in New York, NY

The transition from escalated force to negotiated management tends to place analytic primacy on factors explicitly linked to social movement mobilization and civil disorders. Occurring in parallel to the transition, however, was a major restructuring of the New York Police Department’s (NYPD) approach to crime control coupled with a significant rise and fall of crime between 1960 and the 1990s (Johnson et al., 2000; Zimring, 2012). The policing of protest cannot be decoupled, either conceptually or empirically, from the broader strategies of crime control dominant in a particular locality. When examining the temporal coevolution between crime control and protest control, there is a striking association between the crime drop starting in the 1980s widely heralded by criminologists (e.g. Blumstein and Wallman, 2000) and the shift in dominant strategies of protest policing. Figure 3.1 provides the annual trends in crime rates in New York, NY between 1960 and 2006, based on the Federal Bureau of Investigation’s Uniform Crime Report (UCR) series. There is also a dashed horizontal line differentiating between the periods when escalated force and negotiated management are presumed to be dominant.

Figure 3.1 distinguishes between violent crimes and property crimes, and though the two types of crime differ in scale, their trajectory is roughly comparable.\(^1\) It is clear that starting in 1960 there is a upward shift in both violent and property crimes lasting until the 1980s, at which point we see the beginning of a signifi-

---

\(^1\) Violent crimes include all murders, rapes, robberies, and aggravated assaults while property crimes include burglaries, larcenies, and car thefts.
significant decline in crime rates. In the context of protest control, there is a temporal correspondence between the rise of crime during the escalated force era, and the decline of crime during negotiated management. We would expect then, that as crime rates declined, so too did police aggressivity when dealing with protestors.

This relationship may initially appear to be intuitive but it is difficult to reconcile with the institutional changes in the NYPD that started in the 1980s. Following influential work by Wilson and Kelling (1982) the NYPD adopted an extremely aggressive zero-tolerance approach to crime control coupled with an expansion in hiring and deployment of police officers (McArdle and Erzen, 2001). Consequently,
in the case of New York it is well worth questioning whether the relative permissiveness of negotiated management ever took root, and if it did not, how might we understand the changes in the NYPD’s approaches to protest control?

3.2.3 Policy Spillover: Spatial Privatization and Crime Control

To account for the seeming contradiction between the theoretical expectations of protest policing scholars with the policy decisions adopted by the NYPD, I develop the concept of *policy spillover*, which refers to latent and perhaps unintentional spillover effects that municipal policies and crime control strategies have on the relative permissiveness or severity of protest policing. Policy spillover emerged from a series of policy decisions starting with the Koch Coalition and adoption of Broken Windows policing strategies that resulted in the development of an approach to crime control that is inconsistent with the central tenants of negotiated management but also distinct from escalated force. Unlike the transition towards a softening of protest control suggested by the negotiated management model, the account I develop below predicts just the opposite. In short, since the 1980s the social space needed for protest conflicted with a decidedly pro-business agenda and intolerance for disruption adopted by each mayoral administration starting with Edward Koch.

3.2.3.1 The Koch Coalition and Spatial Privatization

The Koch Coalition dominated the political scene during the 1980s and provided a conservative reclamation of New York City under the management of Mayor Edward Koch, first elected in 1977. It catered primarily to the interests of the white
middle class, European immigrant groups, real estate investors, and the burgeoning Latino and black middle classes (Mollenkopf, 1994). During the 1970s New York faced a severe fiscal crisis resulting in extensive layoffs of municipal workers coupled with crippling crime rates (Arian et al., 1991). The opportunity to revolutionize the municipal political process was seized by Koch, who ran an unusually conservative campaign based on fiscal responsibility, crime reduction, and increasing capital investment in the city. To do so, Koch relied on the demographic reality that non-Latino whites were not only the largest voting block in city—representing 79% of the electorate in 1969—but they were also vastly overrepresented in their turnout rates for municipal elections (Soffer, 2010). Such residents were seeking aggressive solutions to the social problems plaguing New York. These voting blocks would come to form the basis of the Koch Coalition that launched an era of change in New York, resulting in Koch’s reelection in 1981 and 1985 before being defeated by David Dinkins in 1989.

While a thorough discussion is beyond the scope here (but see Mollenkopf and Castells, 1991; Arian et al., 1991; Mollenkopf, 1994), a number of policies and priorities form the core of the Koch Coalition that have ramifications for the policing of protest. Most centrally, the Koch administration actively forged coalitions with elites from the finance, real estate, and law sectors in hopes of building the commercial infrastructure in Manhattan and elsewhere in the city. This was achieved by prioritizing specific building projects, creating tax incentives, and developing zoning regulations designed to obfuscate the unified planning vision of the Koch administration (Mollenkopf, 1994). While the extensive high-rise development gave rise to conflict with minority groups and labor unions (Arian et al., 1991), it remained unabated and the 1980s was a period of urban restructuring that ultimately abated New York’s economic crisis by the end of Koch’s mayoral tenure.
Given the widespread economic restructuring, gentrification and thickening of private industry a durable legacy of the Koch Coalition was the privatization of public space (Soffer, 2010). The movement towards privatization is not unique to New York and instead part of a broader trend across the United States (Mitchell, 1995, 1997, 2003), however, its timing, scale, and centrality to the Koch administration’s strategy had implications for collective dissent. The emphasis on attracting and retaining private investment, including the privatization of many of New York’s parks, 2 permanently altered the purposes for and accessibility of public space (Zukin, 1995). Privatized space, even when publicly accessible, has fewer constitutional protections than fully public space which serves to limit its accessibility to protestors (McCarthy and McPhail, 2006; Zick, 2009). Privatization also shifts the expectations for the spatial routines appropriate for a particular location, or to borrow the terminology from Lefebvre (1991 [1974], 38–40), it redefines the spatial practice of a location. Privatized space is intended primarily to facilitate business without disruption, putting it at odds with simultaneously providing a location for protestors to challenge or air grievances. As a result, we should expect an increase in police hostility over time as protestors come to encroach on or disrupt spatial practices.

3.2.3.2 Broken Windows and the Diffusion of Disorder

While the Koch Coalition transformed the economic landscape in New York by widespread privatization, dealing with the escalating levels of crime and making New York a city safe for tourism remained a priority for Koch (Mollenkopf and Castells, 1991). The 1970s and 1980s also marked the beginning of a major shift

---

2 It should be noted, however, that many parks where in a state of significant disrepair prior to their privatization and city parks were often used as accessible locales for selling drugs.
in existing law enforcement policies aimed at bringing crime under control, most notably through the adoption of a Broken Windows approach to policing. The essence of Broken Windows came from an article published in *The Atlantic* by Wilson and Kelling (1982). Their central argument was that neighborhoods with visible signs of disorder—including graffiti, homelessness, dilapidated buildings, property crime, or litter—signaled a lack of formal and informal social control to potential criminals resulting in higher levels of more serious crimes such as murder or sexual violence. The solution was to increase foot patrols throughout New York to give residents a sense of security coupled with a zero-tolerance approach to low-level misdemeanor crimes that could contribute to a sense of disorder. Whether Broken Windows is responsible for the decline in crime apparent in Figure 3.1 is the subject of some debate—compare for example Harcourt’s (2001) critical analysis to Zimring’s (2012) optimism—but there is a strong consensus that it permanently transformed the politics and practice of crime control in New York.

As mentioned above, a core component of Broken Windows policing is the control over all forms of disorder. Providing a precise definition of disorder has (not surprisingly) proved to be a thorny issue, both theoretically and in terms of policy formulation (Skogan, 1992), and the standard operating definitions are incredibly vague. In their initial exposition of Broken Windows, Wilson and Kelling (1982, 29) frame disorder in terms of citizen fear of crime, claiming that law enforcement should reduce

\[ \ldots \text{the fear of being bothered by disorderly people. Not violent people,} \]
\[ \text{nor, necessarily, criminals, but disreputable or obstreperous or unpredictable people: panhandlers, drunks, addicts, rowdy teenagers, prostitutes, loiterers, the mentally disturbed.} \]
Similarly, Kelling and Coles (1996, 14) claim disorder broadly refers to “... incivility, boorish and threatening behavior that disturbs life, especially urban life,” which includes the obstruction of public spaces, public drinking, vandalism, or unsolicited car window washers. Such sentiments are impractical to implement as official policy since ‘rowdy teenagers’ are subject to the same constitutional protections as others, and the obstruction of public space can be easily implemented in a discriminatory fashion contingent on who is doing the obstruction. The Broken Windows program, though touted by politicians and police leaders, almost immediately resulted in strong allegations of overly aggressive police conduct (McArdle and Erzen, 2001) particularly for socially disadvantaged groups such as the homeless (Vitale, 2008) and racial minorities (Gelman et al., 2007).

A major implication of Broken Windows for protest is that the police became highly protective of any real or perceived misuse of public space, which would convert a contentious demonstration into a criminal gathering. While social movements scholars have traditionally held a firm theoretical line differentiating between protest tactics and criminal behavior (Oliver, 2008), this is not necessarily true from the perspective of law enforcement. Indeed there is evidence supporting a historical coupling of the two during the civil rights era (McAdam, 1999). The conflation of protest and disorder is also a tenant of the escalated force model and was used to justify curtailing First Amendment expression. Research by Schweingruber (2000) finds that police training manuals regularly frame demonstrations using concepts from mob sociology.3 There is also corroborative evidence specific to New York suggesting that more contemporary protest activity may be viewed as

3 The foundational work in mob sociology is by Le Bon (1897), who argues that crowd dynamics take on a life of their own, forming a unified collective mind that is superordinate to individual preferences. Despite its lingering popularity in the law enforcement community, mob sociology has been wholly discredited in the contemporary literature (for the most thorough critique, see McPhail, 1991).
a form of disorder by law enforcement. Specifically, based on more current ethnographic research of several major New York City protests, Vitale (2005, 2007) concludes that the NYPD came to view their role during protests as a form of crime control rather than the protection of the constitutional right to assemble. This is reflected, for example, in the extensive use of arrests and barricades during the 2004 Republican National Convention (Earl, 2009) or the mobilization against the Iraq war of 2003 (NYCLU, 2003).

3.2.3.3 Policy Spillover and the Ascendancy of Aggressive Protest Policing

The structural and institutional changes in New York discussed above appear to have little direct relevance to protest and the articulation of First Amendment rights. Spatial privatization was primarily intended to grow New York’s economy and end the fiscal crisis of the 1970s. Likewise, in principle Broken Windows was designed to address the urban blight and the major social problems caused by the city’s crime rate. Despite this there is the strong possibility that these policy shifts also had latent effects resulting in a new and more aggressive approach to protest policing, a process I defined as policy spillover. The act of protest within a social space designed to facilitate business, tourism, and investment challenged its primary functions, which was made more contentious due to a conceptual linkage between protest activity and disorder. To the degree that a policy spillover explanation is plausible, we should expect to see more aggressive protest policing taking place beginning in the 1980s and continuing until the present. To examine whether this is the case I now turn to an analysis of over six thousand protest events occurring in New York between 1960 and 2006.
3.3 Data and Methods

To examine trends in the policing of protest in New York, NY between 1960 and 2006, I use two separate sources of data that were merged for this analysis. First, I use a subset of the Dynamics of Collective Action (DoC) database collected by McAdam et al. (2011), which contains all protest events reported in the *New York Times* occurring in the United States that took place between 1960 and 1995. To construct the sample of protest events research assistants were instructed to read all daily issues of the *New York Times* and flag all articles mentioning collective action events taking place in the United States during the analytic period. The pertinent articles were then content coded using a standardized codebook. Regular reliability checks were used to ensure consistent coding, and agreement was consistently at or above 90% agreement. A more thorough discussion of this data and its construction is available in a variety of studies including Soule and Earl (2005) or Rafail et al. (Forthcoming). Since this research examines only events taking place in New York City, I reduced to the national DoC database to only those events occurring within the five boroughs that comprise New York City (i.e. Manhattan, the Bronx, Brooklyn, Queens, and Staten Island).

The second source extends the DoC time-series and uses the New York City subsample of my data on protest in 20 U.S. cities (see Chapter 2 for further discussion). I use full-text searches to locate protest events appearing in the *New York Times* between January 1, 1996 until December 31, 2006. Broad search operators were adopted to identify articles covering protest events, yielding a total over 130,000 candidate articles. Each article was examined to determine its relevancy and all

---

4 The dataset and documentation are freely available on the World Wide Web at: [http://www.stanford.edu/group/collectiveaction/cgi-bin/drupal/](http://www.stanford.edu/group/collectiveaction/cgi-bin/drupal/)

5 Though the full-text searches used for the period between 1996 and 2006 are not directly comparable to full-text reads used for the 1960 to 1995 data collection, I compared the events
articles detailing protests events occurring in New York City were content coded by research assistants. I conducted weekly reliability checks and all variables used in the analysis consistently had at least 90% agreement. The final analytic sample combines both sources of data and ultimately provides a sample of 6,147 protest events.

Before turning to my operationalization and analytic strategy, it is important to discuss the limitations of relying on newspaper data as the sole source of information on protest events. Newspapers do not provide a census or systematic random sample of all the protest events that occur in a particular location. Several studies have consistently demonstrated that newspaper coverage is biased towards more spectacular events that feature violence, conflict, or visually striking tactics, which are considered to be higher interest to readers and subscribers (McCarthy et al., 1996; Barranco and Wisler, 1999; Earl et al., 2004; McCarthy et al., 2008). A second form of bias inherent in newspaper data relates to the geographic proximity of a protest event to the editorial office of a newspaper. Research by Myers and Caniglia (2004) and Ortiz et al. (2005) suggests newspapers have a strong preference for events taking place in the community of the newspaper, likely since such events are judged to be of higher interest to local readers and information gathering costs are lower. The use of the New York Times to elicit a sample of protest events minimizes the latter type of bias, though does not address the over-representation of sensational events. To test the veracity of my results, I followed a technique introduced by Soule and Davenport (2009) that involves first randomly eliminating intense events (e.g. large, contentious events) and then re-estimating the statistical models. The substantive results remained robust even with 30% returned by both for two randomly selected months. The pool of events returned was identical suggesting that the two different data collection strategies will produce generally reliable and consistent collections of event coverage.
of intense events removed, suggesting that analyses discussed below are robust to disproportionately high number of certain types of events.

### 3.3.1 Variables and Operationalization

Similar to other research examining the temporal evolution of protest policing (e.g. Soule and Davenport, 2009; Davenport et al., 2011; Rafail et al., Forthcoming) I use two response variables to examine protest policing: whether any arrests occur or whether the police use any other form of force during the course of the protest event. Both of these variables are operationalized as binary indicators with categories for whether (= 1) or not (= 0) they occur. While the use of independent indicators does not allow for the examination of protest policing styles similar to research by Earl et al. (2003) and Earl and Soule (2006), there is a strong advantage to the operationalization I use here. Specifically, I need not assume that a particular style of protest policing remains constant over the entire analytic period. There are also good reasons to expect that arrests and the use of force will follow different trajectories, making it important to keep them as separate response variables. Even though the transition from escalated force to negotiated management is questionable in the case of New York City, McPhail et al. (1998) do make a compelling case that the police relied more heavily on the use of force in the 1960s compared to arrests. If true, we should expect more aggressive police action at the onset of the civil rights movement. The adoption of broken windows also has more focused implications for arrests (Kelling and Coles, 1996), which we should expect to increase more rapidly relative to the use of force since arrests allow the police to clear ‘disorder’ from the streets more easily.

To measure situational threats, I use three dichotomous indicators of presence
(= 1) or absence (= 0). The first is a measure of civil disobedience, defined as non-
violent forms of resistance such as blocking an intersection or passively resisting
arrests. Next, I use an indicator for events where the protestors engage in or
attempt to damage personal or private property. Last, I use a measure of whether
the protestors use or threaten the police or bystanders with any form of physical
violence. Previous literature has consistently found that high levels of situational
threats are more likely to result in police action (Earl et al., 2003; Earl and Soule,
2006; Davenport, 2007a; Soule and Davenport, 2009; Earl, 2011) and I also expect
to find a positive relationship between situational threats and both arrests and the
police use of force.

I use three measures to capture protest characteristics which are non-behavioral
elements of an event that may still influence police conduct. First I include a
variable for the size of each event. In approximately 45% of protests the media
report(s) did not provide a specific estimate of the event size. However, in these
cases it was possible to use textual descriptions in the articles to place each event
into one of six size categories, and this information was used to impute the missing
cases. To do so, for category c, the missing value was replaced by a number drawn
from a uniform probability distribution, such that $\text{Size}_{\text{miss}} \sim U(c_{\text{min}}, c_{\text{max}})$. Since
there are a small number of very large events that skew the distribution I use the
natural logarithm in the regression analysis. Second, I use a dichotomous indicator
for whether (= 1) or not (= 0) any counterdemonstrators were present during each
event, since two groups with disparate political views might alter how the police
attempt to control the protest site. Third, I use another indicator for whether
(= 1) or not (= 0) a formal social movement organization (SMO) was reported to
be present since they may develop cooperative relationships with the police over
time.
A trichotomous variable is used to measure the central claim of each event. I distinguish between minority rights events, peace demonstrations, and all other claims. Clearly this is only a small number of the total number of possible claims available, however the police were disproportionately aggressive towards protests where minority rights and peace were the central claim (e.g. McAdam, 1983; Donner, 1990; Cunningham, 2003a,b; Davenport et al., 2011). There is also evidence that the New York Times gave preferential treatment to the civil rights movement (Rafail et al., 2012), making the inclusion of the minority rights events important controls for potential selection bias. The category for ‘all other claims’ is used as the reference category in the statistical analysis below.

Two annual measures of crime are used to assess whether particularly high or low rates of violent crime or property crime have an influence on protest policing. Similar to Figure 3.1 above, these are drawn from the Uniform Crime Reports that the NYPD reports to the FBI. I take the natural logarithm of each variable to account for particularly high values. To avoid errors of causal ordering the measures of crime are lagged one year. If the political construction of crime had an effect on the NYPD’s approach to protest policing, we should expect that crime rates will have little direct impact on the policing of protest.

I use a series of dichotomous variables for mayoral periods to examine temporal trends in protest generally and potential effects of policy spillover specifically. Previous research attempted to model time in a variety of ways ranging from using a dichotomous distinction between events occurring before or after 1969 (Soule and Davenport, 2009) to non-parametric estimation (Davenport et al., 2011; Rafail et al., Forthcoming). As noted above, policy spillover operates through latent effects of larger institutional commitments and policies, making it quite difficult to measure directly.
One option would be to include indicators for each police commissioner over the analytic period, however, this is limited for several reasons. First, the Police Commissioner is appointed by the Mayor of New York and presumably each mayor is likely to select a commissioner sharing his vision of crime control. Second, a total of 15 commissioners were appointed between 1960 and 2006, suggesting that the changes in police commissioners are so frequent that they are unlikely to result in major policy shifts. Third, policy spillover may arise from factors other than crime control strategies making police commissioners a rougher proxy than mayoral period. The use of several periods is more defensible, since multiple indicators allow one to assess whether any shifts in protest policing remained consistent over time as well as evaluate relative differences in police permissiveness during each mayoral administration. If the police did adopt a more aggressive approach to protest control following the Koch Coalition, we should expect a consistent increase in the likelihood of arrests and the use of force for the mayoral periods beginning with Koch and continuing until the Bloomberg administration. In the regression analysis that follows I use events occurring during the Beame administration as the reference category since this is immediately prior to the election of Koch, and will therefore allow more direct comparison of how protest policing changed before and after the pivotal period of the 1980s.

3.3.2 Analytic Strategy

I use Bayesian semiparametric multilevel logistic regression to predict arrests and other forms of police force. To account for the temporal structure of the data, I treat protest events as nested in years and therefore use two level models. There are several advantages to this analytic strategy. First, newspaper data is not drawn
from a random sample, which makes the \( p \)-values used in frequentist inference difficult to interpret (despite their ubiquity). In contrast, Bayesian estimation allows for direct probabilistic assessments for each parameter in the model, given the data and priors, allowing for a more intuitive way to make statistical inference. Second, there is a growing body of literature suggesting that in the case of generalized mixed models, Bayesian approaches have better estimation properties relative to Penalized Quasi-Likelihood or other estimators based on adaptive quadrature (Diaz, 2007; McCulloch and Searle, 2001; Neuhaus and Segal, 1997; Browne and Draper, 2006). This is particularly the case for the estimation of uncertainty around variances for random effects parameters (Gelman and Hill, 2007). I use uninformative priors on all parameters, and each model consists of three parallel Markov Chains run for 100,000 iterations.

Rather than beginning with the extremely strong assumption that a single parametric effect is sufficient for continuous explanatory variables, I start by testing for complex, non-linear relationships. For each response variable, I fit a model with each continuous variable entered non-parametric function. The results were examined visually to assess whether the relationships departed from linearity. For the model of arrests a linear approximation was appropriate for all variables, but for the analysis of the police use of force, the effect for event size was clearly non-linear. To estimate the non-parametric smooth for event size I use twenty evenly spaced quantile-based knot points, which is sufficient to provide a valid representation of underlying non-linear relationships (Ruppert et al., 2003; Crainiceanu et al., 2005; Keele, 2008).

Since arrests and the police use force are operationalized as dichotomies, I
assume that

\begin{align*}
\text{Arrests} & \sim BR(p) \\
\text{Use of Force} & \sim BR(p)
\end{align*}

In the general case it follows that

\[ p_i = \frac{1}{1 + \exp^{-\eta_i}}, \]

allowing the model to be written as

\[ \eta_i = \gamma_{00} + \mathbf{x}_\text{Sit. Threats}' \gamma + \mathbf{x}_\text{Prot. Char.}' \gamma + \mathbf{x}_\text{Claims}' \gamma. \]

For the models estimating the police use of force a non-parametric smooth for event size, \( f(\text{Size}) \) is substituted into the equation above, which can be expanded to

\[ f(\text{Size}) = \text{Size} \times \gamma + \mathbf{z}' \mathbf{b}, \]

where \( \mathbf{b} \) is a vector of Gaussian random effects and \( \mathbf{z}' \) is a row from the \( n \times 20 \) matrix of knot deviations \( \mathbf{Z} \), defined as

\[
\mathbf{Z} =
\begin{bmatrix}
|\text{Size}_i - \kappa_1|^3 & \ldots & |\text{Size}_i - \kappa_k|^3 \\
\vdots & \ddots & \vdots \\
|\text{Size}_n - \kappa_1|^3 & \ldots & |\text{Size}_n - \kappa_k|^3
\end{bmatrix}.
\]

Finally, I include a random effect parameter and several contextual predictors for the intercept, \( \gamma_{00} \), where for year \( j \) the intercept equation can be written as

\[ \gamma_{0j} = \gamma_{01} + \mathbf{x}_\text{Crime}' \gamma + \mathbf{x}_\text{Mayors}' \gamma + u_j, \]
and \( u \sim \mathcal{N}(0, \sigma_u^2) \).

### 3.4 Results

To establish a baseline of police conduct, Figure 3.2 contains the percentage of events with either arrests or the police use of force annually since 1960. The results in the figure are striking and difficult to reconcile with the presumed transition from escalated force to negotiated management. The annual percentage of events with arrests remains relatively consistent between 1960 and the start of the 1980s and there is a small decline in the percentage of events with arrests during this time period. Rather than the expected increase in permissiveness in protest control since the 1980s, we see a sizable increase in the use of arrests and a reversal of the downward trajectory in the use of force. While there is quite a bit of variability in the frequency that each tactic is used, the overall upward tendency is quite pronounced. Instead, the directionality of the trends is more consistent with the implementation of an aggressive, zero-tolerance approach to crime control following the election of Edward Koch that had a lasting legacy.

While the results in Figure 3.2 suggest that protest policing has become more aggressive since 1960, it is possible that this is the product of a proportionate rise in contentious protestor tactics during this period since threatening behavior is a strong predictors of police action (Earl et al., 2003; Earl and Soule, 2006; Soule and Davenport, 2009). It is therefore also instructive to visualize the temporal trends in protestor tactics over the same period. Figure 3.3 provides the annual percentages for situational threats, including trends in civil disobedience, property damage, and violence between 1960 and 2006. There is considerable volatility in the rates of civil disobedience over time, but overall there is a downward trend its
frequency. Likewise, events with property damage are also notably less common over time, and after 2001, consistently occur at less than 5% of protests. There is also an overall decline in the rates of violence committed by protestors, though there are notable peaks in around 1970 and again in 1990. Taken as a whole, there is little systematic evidence to suggest that protests in New York more regularly used contentious tactics over time, and instead, there has been a notable decline in situational threats.

Taken together Figure 3.2 and Figure 3.3 point to a startling contrast: as protestors became less contentious the police became more aggressive.
away from contentious tactics is consistent with the suggestions that protest groups have become institutionalized across the United States (McCarthy and McPhail, 1998; Meyer and Tarrow, 1998), but inconsistent with widely held expectations by protest policing scholars. The more coercive approach to protest policing starting in the 1980s are consistent with the implementation of zero-tolerance approaches to crime control, providing preliminary support for the policy spillover explanation developed above.

Descriptive statistics for the response and predictor variables are provided in Table 3.2. Arrests took place at a total of 17% of the events reported in the
Table 3.2. Descriptive statistics for protest events occurring in New York, NY between 1960 and 2006 ($n=6,147$)

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Response Variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arrests</td>
<td>0.17</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Police Use of Force</td>
<td>0.08</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td><strong>Situational Threats</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Civil Disobedience</td>
<td>0.17</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Property Damage</td>
<td>0.07</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Use of Violence</td>
<td>0.15</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td><strong>Protest Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Event Size (ln)</td>
<td>4.64</td>
<td>2.12</td>
<td>0.00</td>
<td>13.12</td>
</tr>
<tr>
<td>Counterdemonstrators Present</td>
<td>0.06</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>SMO Present</td>
<td>0.46</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td><strong>Claim</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minority Rights</td>
<td>0.21</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Peace</td>
<td>0.11</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td><strong>Crime Rates</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Property Crime Rate (ln)</td>
<td>3.74</td>
<td>0.49</td>
<td>2.60</td>
<td>4.39</td>
</tr>
<tr>
<td>Violent Crime Rate (ln)</td>
<td>2.34</td>
<td>0.77</td>
<td>0.87</td>
<td>3.51</td>
</tr>
<tr>
<td><strong>Mayoral Periods</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wagner, 1960-1965</td>
<td>0.16</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Lindsay, 1966-1973</td>
<td>0.30</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Beame, 1974-1977</td>
<td>0.08</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Koch, 1978-1989</td>
<td>0.17</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Dinkins, 1990-1993</td>
<td>0.08</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Giuliani, 1994-2001</td>
<td>0.13</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Bloomberg, 2002-2006</td>
<td>0.08</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td><strong>Event Year</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>19.26</td>
<td>13.57</td>
<td>0.00</td>
<td>46.00</td>
</tr>
</tbody>
</table>

The New York Times, while other forms of police force were much less common taking place at only 8% of demonstrations. Situational threats are also relatively common at events, with civil disobedience occurring in 17% of cases, followed closely by protestor violence which occurred in 15% of demonstrations. The results for protest characteristics indicate that the average event has approximately 100 participants ($= \exp(4.64)$), though this is highly variable. Counterdemonstrators are only present at 6% of events, but consistent with claims that American social
movements have become professionalized (McCarthy and McPhail, 1998), a formal social movement organization was reported to be present at nearly 50% of events. Nearly one-third of the events had either minority rights or peace as the primary claim and there is nearly twice as many of the former than the latter. The descriptive statistics for logged crime rates show that, as we should expect given Figure 3.1, property crime is much more common than violent crime, though the standard deviation is much higher for violent crimes. Finally, there is a relatively even split between events occurring before and after the policy reformations started by Mayor Edward Koch. A total of 54% of protests in the sample occurred during the Wagner, Lindsay, and Beame administrations, while the remaining 46% took place during the administrations of Koch, Dinkins, Giuliani, and Bloomberg.

The first set of regression analyses in Table 3.3 provides the posterior means and standard deviations for four models predicting the use of arrests. I use a nested model approach, adding a new conceptual group of variables with each subsequent model. The first includes only the covariates measuring situational threats, followed by the inclusion of protest characteristics, then the central claim of the event, and finally I include the annual measures of logged crime rates and the matrix of variables measuring each of the mayoral periods. Starting with Model 1 in Table 3.3, there is a clear increase in the likelihood of arrests at events where either civil disobedience occurs or the protestors engage in violence. The 95% credible intervals remain consistently positive across the different model specifications and this corroborates previous research (e.g. Soule and Davenport, 2009) suggesting that when the protestors use aggressive tactics the police are likely to respond in a similar manner. The exception is the effect of property damage: while the 95% credible interval for property damage overlaps zero in the first specification, it becomes wholly positive once the other measures are added.
Table 3.3. Bayesian multilevel logistic regression estimates of the factors predicting arrests at protest events in New York, NY between 1960 and 2006 (n=6,147)

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Situational Threats</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Civil Disobedience</td>
<td>1.96*</td>
<td>1.99*</td>
<td>2.00*</td>
<td>2.01*</td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.09)</td>
<td>(0.09)</td>
<td>(0.09)</td>
</tr>
<tr>
<td>Property Damage</td>
<td>0.26</td>
<td>0.27*</td>
<td>0.28*</td>
<td>0.29*</td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(0.14)</td>
<td>(0.14)</td>
<td>(0.14)</td>
</tr>
<tr>
<td>Use of Violence</td>
<td>1.96*</td>
<td>1.98*</td>
<td>2.00*</td>
<td>2.02*</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.10)</td>
<td>(0.10)</td>
<td>(0.10)</td>
</tr>
<tr>
<td><strong>Protest Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Event Size (ln)</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td></td>
</tr>
<tr>
<td>Counterdemonstrators</td>
<td>0.27</td>
<td>0.19</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.16)</td>
<td>(0.16)</td>
<td>(0.16)</td>
<td></td>
</tr>
<tr>
<td>SMO Present</td>
<td>0.04</td>
<td>0.01</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(0.08)</td>
<td>(0.08)</td>
<td></td>
</tr>
<tr>
<td><strong>Claim</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minority Rights</td>
<td>0.24*</td>
<td>0.25*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peace</td>
<td>0.64*</td>
<td>0.63*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.13)</td>
<td>(0.13)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Crime Rates</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Property Crime Rate (ln)</td>
<td>-0.24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.68)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Violent Crime Rate (ln)</td>
<td>-0.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.47)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mayoral Periods</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wagner, 1960-1965</td>
<td>0.64</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.68)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lindsay, 1966-1973</td>
<td>0.67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.46)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Koch, 1978-1989</td>
<td>0.99*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.38)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dinkins, 1990-1993</td>
<td>1.03*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.41)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Giuliani, 1994-2001</td>
<td>1.44*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.38)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bloomberg, 2002-2006</td>
<td>1.31*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.52)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\alpha$</td>
<td>-2.63*</td>
<td>-2.83*</td>
<td>-2.91*</td>
<td>-2.67*</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.14)</td>
<td>(0.14)</td>
<td>(0.17)</td>
</tr>
<tr>
<td>Variance, $\sigma^2_\alpha$</td>
<td>0.55</td>
<td>0.54</td>
<td>0.55</td>
<td>0.40</td>
</tr>
</tbody>
</table>

*Note: Reference category is events under Beame. MCMC standard deviations in parentheses. *=95% credible interval does not overlap zero.
The inclusion of the measures for protest characteristics does little to improve the overall fit regardless of the model specification. As a group, these measures have little explanatory power. The posterior means for each of the variables are positive though quite small in the case of event size and social movement organization presence, and there is quite a large amount of variation in the posterior distributions. In all cases, the credible intervals overlap zero and we therefore cannot definitively assign a specific directionality to the relationship with 95%. This pattern of results is quite significant for two reasons. First, several previous studies have found a robust positive association between event size and arrests (e.g. Earl et al., 2003; Earl and Soule, 2006), perhaps since larger events are more threatening to the police, but this relationship does not appear to be supported in New York. Second, there is little evidence to support arguments that the presence of a social movement organization might be privileged by police. Since it is much more straightforward for the police to negotiate and communicate with an SMO, the lack of effect for this variable casts further doubt that central elements of the negotiated management model were adopted by the NYPD.

Once the variables measuring the central claim of the event are added in Model 3, we see that arrests are more likely to occur when the protestors are mobilizing for issues related to minority rights and peace. This extends previous research indicating that the police are inconsistent in their tolerance of protests initiated by minority groups and particularly for African Americans (Davenport et al., 2011). In the case of New York, the positive effect may be a result of a history of hostility between the minority communities and the police that continued long after the decline of the civil rights movement. For example, the 1999 death of Amadou Diallo, an unarmed Guinean immigrant shot by the NYPD launched a campaign where hundreds of African American protestors (including former mayor David
Dinkins) were arrested during daily demonstrations at the NYPD’s headquarters. Similar waves of contention occurred in 1997 after several NYPD officers assaulted and sodomized Abner Louima with a toilet plunger and the NYPD’s involvement in the deaths of Sean Bell in 2006 and Patrick Dorismond in 2000 also sparked significant public order disturbances. In terms of the peace movement, as noted above, the NYPD had a particularly heavy-handed reaction to the Iraq war protests starting in 2003 and beyond (NYCLU, 2003), but the department’s anti-radical unit was also highly involved during the Vietnam war as well (Donner, 1990). It is important to further highlight that these effects have strong implications in inequalities in the articulation of First Amendment rights, particularly since the claim of the protest should have little impact on police action once the behaviors of the protestors are included in the regression model.

The variables measuring annual crime rates add little to the model. In fact, the posterior means for both property crime and violent crime are negative, though the posterior standard deviations are quite large and the 95% credible interval overlaps zero. These estimates are quite meaningful and suggest that patterns of arrests do not appear to be directly influenced by the high or low rates of crime control in a community. In many ways, the lack of systematic relationship between crime control and arrests is quite surprising since high levels of crime may stretch police resources leaving little incentive to shift police resources and labor power towards protest control, while at the same time low levels of crime might allow the police to mobilize higher numbers of officers during events and monitor events more consistently.

The lack of effects for crime rates are more intuitive when viewed in the context of the estimates across the different mayoral periods. There is a clear threshold where events occurring after Koch’s election are policed more aggressively than
events taking place prior to it. Indeed compared to protests occurring during the Beame administration, the likelihood of arrests is largely equivalent during the administrations of Wagner and Lindsay, and then remain consistently high for the rest of the analytic period. Starting with the election of Koch, there appears to be an increase in the likelihood of arrests that has held constant until 2006. The overall pattern from the contextual variables is that crime rates themselves have little influence on influence police conduct. Instead the political construction of crime control and the institutional policies and priorities have resulted in a higher likelihood of arrests over time rather than a softening of police tactics that is widely expected by scholars. The pattern of parameters in Table 3.3 is quite consistent with a policy spillover explanation, but cannot be reconciled with the transition from escalated force to negotiated management.

The regression results for the police use of force are summarized in Table 3.4. As before I use the posterior means and standard deviations to summarize individual components of the joint posterior density. The overall pattern of results for situational threats is consistent with the estimates for arrests. The 95% credible intervals for both civil disobedience and property damage are uniformly positive in Model 1 and this holds even when additional covariates are added to the model. Unlike the models predicting arrests, however, property damage does not increase the likelihood that the police will use force and indeed by Model 4 has a negative posterior mean.

The parametric effects for the presence of counterdemonstrators and social movement organizations added in Model 2 are also consistent with the effects for arrests and once again have little effect on the likelihood that the police will use force in the full model specification. The non-parametric smooth from Model 4 is provided in Figure 3.4 and does improve the explanatory power of the model. The
Table 3.4. Bayesian semiparametric multilevel logistic regression estimates of the factors predicting police use of force at protest events in New York, NY between 1960 and 2006 \((n=6,147)\)

<table>
<thead>
<tr>
<th>Situational Threats</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Civil Disobedience</td>
<td>1.39*</td>
<td>1.55*</td>
<td>1.55*</td>
<td>1.58*</td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
<td>(0.13)</td>
<td>(0.13)</td>
<td>(0.13)</td>
</tr>
<tr>
<td>Property Damage</td>
<td>0.01</td>
<td>-0.06</td>
<td>-0.07</td>
<td>-0.06</td>
</tr>
<tr>
<td></td>
<td>(0.16)</td>
<td>(0.18)</td>
<td>(0.17)</td>
<td>(0.18)</td>
</tr>
<tr>
<td>Use of Violence</td>
<td>2.68*</td>
<td>2.89*</td>
<td>2.92*</td>
<td>2.97*</td>
</tr>
<tr>
<td></td>
<td>(0.13)</td>
<td>(0.14)</td>
<td>(0.14)</td>
<td>(0.14)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Protest Characteristics</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Event Size (ln)</td>
<td>† †</td>
<td></td>
<td></td>
<td>See Fig. 3.4</td>
</tr>
<tr>
<td>Counterdemonstrators Present</td>
<td>0.39*</td>
<td>0.31</td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.19)</td>
<td>(0.19)</td>
<td>(0.20)</td>
<td></td>
</tr>
<tr>
<td>SMO Present</td>
<td>-0.05</td>
<td>-0.11</td>
<td>-0.10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
<td>(0.12)</td>
<td>(0.12)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Claim</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Minority Rights</td>
<td>0.51*</td>
<td>0.47*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(0.14)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peace</td>
<td>0.80*</td>
<td>0.76*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.17)</td>
<td>(0.17)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Crime Rates</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Property Crime Rate (ln)(_{t-1})</td>
<td>1.44</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.95)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Violent Crime Rate (ln)(_{t-1})</td>
<td>-0.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.65)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mayoral Periods</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Wagner, 1960-1965</td>
<td>2.01*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.94)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lindsay, 1966-1973</td>
<td>0.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.65)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Koch, 1978-1989</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.57)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dinkins, 1990-1993</td>
<td>1.13*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.50)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Giuliani, 1994-2001</td>
<td>1.69*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.55)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bloomberg, 2002-2006</td>
<td>2.03*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.73)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Intercept                                  | -4.00*      | -3.62       | -4.82*      | -7.79       |
|                                            | (0.17)      | (2.56)      | (1.65)      | (4.98)      |

Variance, \(\sigma^2_o\)                     | 0.88        | 0.81        | 0.78        | 0.56        |

Note: Reference category is events under Beame. MCMC standard deviations in parentheses. *=95% credible interval does not overlap zero. †=Smooths available upon request.
Figure 3.4. Nonparametric smooth of event size from a Bayesian semiparametric multilevel logistic regression model estimating the police use of force at protest events in New York, NY ($n=6,147$)

The solid line in the figure represents the posterior mean of the effect of logged event size across the span of the variable, while the dashed lines are used to show the upper and lower boundaries of the 95% credible intervals. There is a clear increase in the likelihood of police force as event size grows, however the scale of the change remains relatively constant between a logged event size of 5 ($\approx \exp(150)$) and 9 ($\approx \exp(8000)$) before starting to rise again for the very large events. There is also a considerable widening in the precision for the largest events likely due to a combination of their rarity as well as a lack of consistent police strategies when they are vastly outnumbered by protestors. Take as a whole, the smooth in Figure 3.4
provides corroborating evidence that large events may threaten the police (e.g. Earl et al., 2003; Soule and Davenport, 2009).

The 95% credible intervals from the claims variables measuring minority rights and peace protests that are introduced in Model 3 are uniformly positive. This is further evidence that the police do not operate with an even hand when it comes to the central claim of the event. As noted above, in the case of minority rights, the high levels of force may be a product of negative ritualization where the police are more to rely on proactive strategies given a long history of hostile interactions with minority groups, which has been seen in previous studies (Rafail, 2010). In the case of peace protests, the pattern of results should also be unsurprising given the NYPD’s history of widespread covert and overt repression of the movement (Donner, 1990). The consistency in directionality and relative scale of the regression estimates strongly suggests that the differential amount of police tolerance in Table 3.3 and Table 3.4 is not an aberration, but instead a robust predictor of differential police action for wholly non-behavioral reasons.

Turning to the contextual variables that are introduced in Model 4, there does not appear to be a systematic relationship between the police use of force and either property crime rates or violent crime rates. The posterior mean for property crimes is positive, however the 95% credible interval overlaps zero making directional claims difficult. The results are quite interesting and suggest that even though the analytic period saw the ascendancy and decline of crime in New York, previous crime rates do not appear to be directly related to force. This pattern of estimates is particularly noteworthy given the adoption of extremely strict zero-tolerance policing strategies under Broken Windows, and provides further evidence that the observed rates of crime are perhaps less important than the political problematization of crime.
The estimates for each mayoral period for the use of force is somewhat different compared to the posterior distributions for arrests, which was expected. There is evidence that force was common during the beginning of the analytic period, and, relative to events taking place during Beame’s tenure as mayor between 1974 and 1977, the police are consistently more likely to use force during Wagner’s administration (1960-1965). This period also marks the beginning of the 1960s wave of contention that was met with a considerable amount of police violence, so these results should be expected. The disproportionately high use of force does not last and the effect disappears with Lindsay’s election in 1966. It is also noteworthy that the police rates of force under Koch are statistically indistinguishable from those taking place under Beame. For the use of force then, there appears to be slower rates of policy spillover. A possible explanation for this is that arrests quickly became the staple police response to protest after the adoption of broken windows, however the highly aggressive strategies resulted in an isomorphic diffusion justifying the use of force as well. Despite the slight difference during the Koch administration, starting in 1990 with the Dinkins administration the use of force becomes more likely and the posterior means continues to grow larger over time, which is consistent with the use of arrests. The general pattern of results is again quite inconsistent with a purported transition from escalated force to negotiated management. Similar to arrests, these results are consistent with the policy spillover account, where the police became less permissive to protest shortly following the push toward spatial privatization and Broken Windows policing strategies.
3.5 Discussion and Conclusion

There are several implications of this study for not only for protest policing in New York but also more generally. Above all else, this research provides strong evidence using two separate measures of police conduct that the widely expected transition from escalated force to negotiated management almost certainly did not occur in New York. Instead, there was a sharp increase in police aggressivity starting in the 1980s—widely held to be the beginning of the negotiated management era—that continued until the end of my analytic period in 2006. To account for the surprising reversal in expected protest policing practices, I developed the concept of policy spillover. Starting with the election of Mayor Edward Koch in 1979, New York City underwent a series of major social and institutional transformations. I focused first on the widespread spatial privatization of the city which was adopted to spur investment, bolster tourism, and increase real estate development in hopes of ending the severe fiscal crisis Koch inherited. The consequence of this bundle of policies was the redefinition of the purposes and accessibility of formerly public space, which previous work suggests often has negative consequences for social movement mobilization (McCarthy and McPhail, 2006; Zick, 2009). Second, in hopes of reducing the growing levels of crime in the city, the NYPD adopted a strict, zero-tolerance approach to crime control, based on the Broken Windows philosophy outlined by Wilson and Kelling (1982). The eventual result of this policy transformation appears to be that protest, and especially contentious protest, came to be treated as disorder, relegating it to a form of potentially criminal behavior rather than legitimate First Amendment expression. Thus, even though neither spatial privatization nor the adoption of Broken Windows policies were intended to have direct consequences for protest they soon ushered in an approach
to protest management that was less permissive and more antagonistic. The results from the statistical analysis here largely confirm these expectations: roughly around the beginning of the election of Koch there was a corresponding and persistent increase in the likelihood that the police used force or made arrests at protest events, even as the negotiated management model was supposedly in effect.

The remaining question is how and whether the results presented here might generalize to other cities. McPhail et al. (1998) are careful to caveat their transitional account by recognizing an uneven diffusion of negotiated management practices to locations outside of Washington, D.C. New York’s social problems and their political solutions are clearly unique and likely do not extend elsewhere. However, the Broken Windows approach to policing became standard in many large cities across the United States (Vitale, 2008) and spatial privatization is increasingly becoming the norm in urban America (Mitchell, 1997, 2003), suggesting that policy spillover may have occurred in other locations as well. It is also almost certainly the case that other municipalities might have experienced as similar growth in aggressive protest policing. If true, perhaps scholarly attention should focus on identifying clusters of municipal policies that have variable impacts on protest policing, which would test and further refine the policy spillover account developed here. Future research and analysis of how protest policing has changed in Washington D.C. would be particularly helpful in this regard.

The results presented here are also challenge the growing tendency to claim that threat rather than context are the most important predictors of action (Earl, 2011). The period between 1960 and the mid 1970s had particularly contentious wave of protest without a corresponding increase in the use of force or arrests. During the early period of John Lindsay’s administration, for example, there was a particularly hostile transit strike marking the beginning of a period of significant
protest and upheaval in New York (see Arian et al., 1991; Caliendo, 2010), yet even here there does not appear to be any evidence indicating that the police became increasingly hostile to protest and protestors. Indeed, focusing on the threatening behaviors of the protestors alone loses the spatial and political contexts where protest occurs. The results here make it clear that not only is context a fundamental component to protest policing practices, but that excluding it serves to weaken and oversimplify existing theoretical understandings through overstating the importance of situational threats that may arise during a protest.

Finally, this research also raises several questions about how, for instance, Supreme Court decisions related to protest trickle down to the municipal level, interact with preexisting public order management strategies, and how such decisions are interpreted and implemented in daily operations. Along with widespread federal training of municipal police departments, McPhail et al. (1998) identify several national policies that almost certainly had some sort of impact on existing approaches to crime control. Unfortunately studies examining the diffusion of court decisions have been essentially non-existent in the protest policing literature. More attention paid to this important topic will help to shed more light on the balance between meso and macro contexts.
4.1 Introduction

Understanding the factors shaping state repressive action is generally recognized as a fundamental component of social movement theory. A robust body of literature has developed examining the interactions between western democratic governments and protest groups (della Porta, 1995; della Porta and Reiter, 1998). While such research has pointed to several key mechanisms explaining state decisions to engage in repression, the majority of scholars have emphasized overt forms of state control, particularly police reactions to protest such as police presence, arrests, or the use of force (McPhail et al., 1998; Earl et al., 2003). It is widely contended that the
behavioral threats posed by the protestors to the police are the best predictors of police action (Davenport, 2000; Earl and Soule, 2006; Davenport, 2007a; Soule and Davenport, 2009), though other research has begun challenging this relationship (Rafail, 2010).

Despite the emphasis on the overt behavior of state actors, most scholars recognize that repression is multifaceted and overt repression is far from an exhaustive enumeration of the tactics available to governments and law enforcement. It has long been established that covert repression can have a dramatic impact on social movement organization, behavior, and outcomes (Marx, 1974). Though a number of scholars have examined covert repression, research focusing on surveillance, infiltration, and subversive activity by state actors remains relatively rare. This lack of attention is generally for good reason: the nature of covert repression makes its use a closely guarded secret and evidence of covert repressive activity often does not come to light for many years (if at all). For this reason, the majority of research on covert repression draws from declassified documents written during the 1960s wave of protest, and focuses on movements such as the new left, the American Indian Movement, black nationalism, and white supremacy (Carley, 1997; Churchill and Vander Wall, 2002; Cunningham, 2003a,b, 2004; Davenport, 2005). Contemporary patterns of surveillance are largely a theoretical black box, though McCarthy et al. (Forthcoming) note anecdotally that covert repression appears to be on the rise across the United States and potentially other western democracies. It is largely an open question whether campaigns of covert repression from the 1960s and 1970s have remained largely consistent with the tactics used for contemporary practices. Given the often devastating impact such practices can have on civil liberties, more research is required on this important topic.

With these considerations in mind, this research focuses on the factors shap-
ing patterns of contemporary covert repression. By doing so I contribute to the literature on state repression, both covert and overt, in several ways. First, I examine contemporary trends in covert repression using a database of over 400 social movement organizations (SMOs) that were surveilled between 2009 and 2010. Second, unlike much of the other research on covert repression that focuses on a small handful of movements or SMOs (e.g. Cunningham, 2003a; Davenport, 2005), I compare surveillance patterns across several different movements and hundreds of organizations. Finally, as Cunningham (2004) demonstrates, covert repression is heavily influenced by the organizational structure of the repressive actor(s). I extend this line of argumentation by outlining the importance of organizational demands and the logic of target selection to covert repression. I then draw on the restructuring of intelligence gathering after the September 11, 2001 terrorist attacks, and emphasize how the current organizational and legal environments considerably widened the scope and frequency of covert repression in the post-9/11 era.

4.2 The Covert Repression of Social Movements

There is a long history of scholarly attention to state repression in social movements research, ranging from Tilly’s (1978) classic work From Mobilization to Revolution to contemporary studies of protest policing. In a review of the literature, Davenport (2007a) conceptualizes repression as any violation of First Amendment-type rights by government actors (e.g. freedom of assembly, association, or peaceful protest). He goes on to propose the “law of coercive responsiveness,” which references the empirical regularity that when a social movement engages in behavior that threatens the state or its institutions, it is likely to provoke a coercive response.
State repression can encompass a wide variety of tactics including overt actions such as genocide, arrests, or other forms of violence as well as covert actions (i.e. undercover agents or surveillance). From this wide variety of options the repertoire of repressive activity used by a state is largely rooted in the popular expectations of state conduct, legal precedents, and institutional norms held by law enforcement organizations (McPhail and McCarthy, 2005). For this reason, western democratic governments rely on comparatively benign responses when faced with threatening challengers, at least relative to more autocratic states that may not hesitate to use lethal force.

Covert repression generally refers to the use of repressive action by the state that is unobservable to social movement participants or the public (Earl, 2003). Covert repression draws from a variety of tactics including supervision, infiltration, disruption, or taking steps intended to dispel or neutralize a social movement (examples are provided in Donner, 1990; Stotik et al., 1994; Churchill and Vander Wall, 2002; Cunningham, 2003a; Irons, 2006; Boykoff, 2007). In contrast overt repression refers to observable tactics used by the state to control or neutralize social movements (Earl, 2003). Like covert repression there are a wide variety of overt repressive tactics available to the state, though due to their visibility such tactics are more heavily tempered by existing political and social arrangements (della Porta, 1995). Overt and covert repression are highly interrelated and are often used in conjunction. For example, the state could instruct agent provocateurs to use violence during a protest event which can justify the police in making arrests. The information gathered from those who are arrested can then be used to heighten or sharpen existing covert repression.

Despite their apparent similarity it is useful to conceptually decouple certain elements influencing patterns of covert repression from those affecting overt re-
pression. Though both types of repression fall along a single continuum, overt and covert repression can differ quite dramatically in their tactics and aims. For example, forms of overt repression such as protest policing are primarily situational and oriented towards handling specific events. Police departments create and implement general organizational strategies for dealing with protests to be sure (McPhail et al., 1998; McCarthy and McPhail, 1998; McPhail and McCarthy, 2005), yet even the most heavily negotiated demonstration cannot be fully choreographed. Police then necessarily adopt a mixture of proactive and reactive responses when faced with a particular situation on the ground. In contrast, covert repression is often more targeted, long-term, and goal oriented. Government agencies use purposive logic to select targets deemed worthy of covert repression with an evolving agenda (i.e. intelligence gathering, infiltration, movement dissolution). Such decisions do not follow rationalist analyses of perceived threat that are commonly invoked in studies of overt repression (Cunningham, 2003a). 1 I now turn to a more detailed discussion of two key mechanisms of covert repression that are distinct from overt repression. A more integrative discussion of the relationship between covert and overt repression is provided in the concluding section of this study.

4.2.1 Key Mechanisms of Covert Repression

To develop a theoretical account of covert repression I propose two mechanisms that aid in differentiating between covert and overt state actions. I first suggest that covert repression is heavily conditioned by the organizational demands of the repressive actor. Second, I argue that the logic of target selection for covert

---

1 By rationalist analyses I refer to arguments suggesting a unidirectional relationship where more threatening social movement actors are met with (roughly) proportionately coercive government responses. Examples of such arguments can be found in Lichbach (1984) or Moore (2000) though many scholars make similar arguments implicitly.
repression is contingent on temporally variable political processes that legitimize the repression of certain social movement actors.

4.2.1.1 Organizational Demands

While covert and overt repression are both conditioned by the organizational demands of the repressive agent, it is particularly the case for the former. Unlike overt forms of repression where visible benchmarks such as public order maintenance can be used to evaluate effectiveness, covert repression is kept secret and its successes and failures must be justified internally. The expenditure of scarce organizational resources on surveillance programs demands that units responsible for such operations continually justify their existence by pointing to previous successes. This creates incentives to gradually expand the scope of groups worthy of surveillance and repression in hopes of assuring a steady or growing stream of funding, technology, and labor power. The need for covert surveillance programs can then easily become self-justifying (Marx, 1974). Widening the scope of which groups constitute legitimate targets to repress in order to meet organizational demands is well documented historically. From Solzhenitsyn’s (2007 [1973]) analysis of the growth of Gulags in the Soviet Union to the exponential growth in the suspected communist sympathizers during McCarthyism (Schrecker, 1998), it is clear that covert repression programs facilitate the creation of a fuzzy set of challengers worthy of oppression. The boundaries of this group are then gradually expanded over time. Such a process becomes further entrenched as organizational members defend their own interests in organizational maintenance (Michels, 1966).
4.2.1.2 The Logic of Target Selection

A significant proportion of overt repression is necessarily a reaction to social movements staging demonstrations or other public displays of collective action (e.g. Earl et al., 2003; Soule and Davenport, 2009), at least in western democratic states. Covert repression has a logic of target selection that more heavily focuses on the political identity or claims of a social movement (Davenport, 2005) and therefore generally follows more proactive and purposive strategy. A consequence of this is that groups or movements are often surveilled or infiltrated when they pose little threat to public order or national security. Cunningham’s (2003a) analysis of FBI surveillance of new left organizations, for example, notes several instances of the high ranking agents from the main DC branch ordering local FBI field offices to target new left organizations that were largely inactive or otherwise of little threat.

A limitation of previous research is that many studies of covert repression have focused on a small number of groups or movements. This has reduced scholarly understanding of why particular movements are selected as appropriate targets for covert repression over others. A key explanatory factor of the selection procedure is the political marginalization of a movement by the state, other elites, or public opinion. By political marginalization I refer to efforts by established political or institutional actors to dismiss or trivialize a movement’s claims, the encouragement of suspicion directed at participants or sympathizers, or other efforts to cast a movement as a political Other. Marginalization can be the result of engaging in particularly contentious mobilization, having a political identity that is fundamentally incompatible with existing political and social arrangements, or more commonly a mixture of the two. Previous research by Gamson (1975) indicates that the state is more likely to repress weak groups since there is a much higher
chance of success. The weakness argument has been examined extensively and few studies provide support direct support for key elements of the position (Earl et al., 2003; Earl, 2006; Earl and Soule, 2006). A potential reason for this might be the conflation of the weakness of a social movement with the degree to which it is politically marginalized. It is quite possible for a group to have a tremendous amount of resources while being politically marginalized (e.g. the marriage equality movement). In this case, the empirical record is much more supportive of an association between either covert repression (Cunningham, 2003a; Davenport, 2005) or overt repression (Stockdill, 2003; Rafail, 2010) and political marginalization.

When it comes to target selection, then, the more a social movement or activist group is politically marginalized by repressive actors the higher the risk of covert repression. There are strong historical precedents to support this claim such as the red scare of the McCarthy era where the political will among elites facilitated the suppression of (perceived) communist activity (Gibson, 1988). Much the same could be said for the covert repression of anarchists after the World Trade Organization meetings of 1999 in Seattle and perhaps the global justice movement as a whole. Directing campaigns of covert repression towards marginalized movements or groups also provides a natural cover if covert repression is exposed, since it is easier to justify actions against such groups to the general public.

4.2.2 From Subversion to Information Gathering?

To concretize and contextualize the impact of organizational demands and target selection in terms of current uses of covert repression, it is useful to place them in historical context. My central argument is that the dominant purpose of contemporary programs of covert repression focus on intelligence gathering rather than
suppression, infiltration, or active attempts at disruption. This is facilitated by the centralization of federal law enforcement after 9/11, the rapid expansion of provisions allowing the government to engage in intelligence gathering, and the legal and discursive linkage of social movement mobilization and domestic terrorism (particularly for the animal rights and environmental movements). To demonstrate this I begin with a brief description of the Federal Bureau of Investigation’s Counter Intelligence Programs (COINTELPRO) and then turn to changes in organizational demands and target selection since the 9/11 terrorist attacks.

4.2.2.1 A Brief Overview of COINTELPRO

Beginning in 1956 and continuing throughout the 1960s, COINTELPROs targeted the new left, racist right (particularly the KKK), communists, and other politically marginalized movements (Churchill and Vander Wall, 2002; Cunningham, 2003a,b). During its heyday, COINTELPROs and similar programs by state and local law enforcement agencies were well resourced and extensively used. Donner (1990, p. 82) estimates that by 1963 over 300,000 local law enforcement officers had some degree of involvement in covert repression programs along with between 6,000 and 7,000 federal agents. It is largely agreed that COINTELPRO and other campaigns of covert repression routinely violated the civil liberties of activists and had stated goals of disrupting or annihilating groups deemed overly subversive by J. Edgar Hoover and other high-ranking FBI officials (Cunningham, 2004). Schultz and Schultz (1989, 2001) provide extensive activist accounts documenting dozens of examples of civil rights violations.

The COINTELPRO campaign took a much more active approach to covert repression and was quite successful. Several scholars have linked COINTELPRO to the decline and eventual dissolution of groups such as the American Indian Move-
ment and the Black Panther Party (Carley, 1997; Churchill and Vander Wall, 2002). Cunningham and Noakes (2008, pp. 181-182) use such activities to differentiate between counterintelligence and intelligence. By the former, they refer to active campaigns of covert repression designed to limit a target’s ability to mobilize or increase the likelihood of illegal activities. Intelligence, in contrast, is a more passive approach primarily concerned with the collection of information about a target’s previous and/or past behaviors. While programs such as COINTELPRO were clearly dominated by counterintelligence, contemporary covert repression is more firmly rooted in a logic of intelligence gathering.

4.2.2.2 Intelligence Gathering Reformation and 9/11

The aftermath of the terrorist attacks of September 11, 2011, propelled new changes in the structure of law enforcement agencies. Several federal agencies such as the United States Immigration and Naturalization Service and Federal Emergency Management Agency were folded into the newly created Department of Homeland Security (DHS). Though imperfect in its execution, the DHS placed considerable emphasis on the intelligence gathering through the centralized control over disparate agencies (Perrow, 2006). The 9/11 attacks were generally framed in terms of intelligence failures, and the DHS was given a budget of tens of billions of dollars with the expectation that it would prevent pending or future terrorist threats.

Along with changes in the organizational structure of law enforcement the post-9/11 era was also witness to significant changes in legislation concerning intelligence gathering, greatly simplifying the task of the DHS and other law enforcement organizations. Of central importance was the October, 2001 passage of the Uniting and Strengthening America by Providing Appropriate Tools Re-
quired to Intercept and Obstruct Terrorism Act (USA PATRIOT Act). The USA PATRIOT Act introduced several provisions that fundamentally altered how and why the government could collect intelligence. Rackow’s (2001) analysis of the USA PATRIOT Act suggests that three provisions were particularly subject to broad interpretation. First, the bill allows for the surveillance of those suspected of domestic criminal activity without the establishment of probable cause; second it allowed the use of roving wiretaps to monitor the conversations of nonsuspects with minimal regard to individual privacy; third, the USA PATRIOT Act could be used to “...discourage political dissent by including the activities of unpopular political organizations within the newly created definition of “domestic terrorism.”” (Rackow, 2001, p.1653).

It is almost uniformly agreed upon by scholars that such provisions increased the scope and scale of intelligence gathering since the USA PATRIOT Act’s passage. A wide variety of legal opinions have been published challenging the constitutionality of the bill (e.g. Whitehead and Aden, 2001; Mell, 2002; Thistle, 2008), particularly its potential chilling effect on the right to lawfully assemble (Saito, 2002; Fisher, 2004). This is not surprising given the broad definition of threat adopted by the DHS and other governmental or law enforcement agencies. In short, the political climate following the 9/11 attacks created intense organizational demands of the DHS that revolved around aggressively monitoring any activities that could be construed as terroristic behavior, both domestically and internationally.

4.2.2.3 Social Movement Mobilization and Domestic Terrorism

Given the mandate of the DHS and law enforcement agencies to eliminate any potential terrorist activity, a more more expansive logic of target selection emerged
and was widely adopted. The language of the USA PATRIOT Act was such that domestic terrorism “...involve[s] acts dangerous to human life that are... in violation of the criminal laws of the United States or of any state” that might appear to “influence the policy of a government by intimidation or coercion,” “intimidate or coerce a civilian population,” or “affect the conduct of a government by mass destruction, assassination, or kidnapping.” (USA PATRIOT Act §802, cited in Saito, 2002). This broadened the scope of what constitutes domestic terrorism and allowed the inclusion of common protest tactics such as civil disobedience, blocking intersections, or failing to obey a police officer (Chang, 2001; Saito, 2002).

The extent to which law enforcement agencies linked social movement mobilization with the USA PATRIOT Act is unclear, though following the 9/11 attacks several anecdotal accounts of activist intimidation emerged that were justified in terms of intelligence gathering. Fisher (2004) provides several examples of federal and municipal law enforcement conducting interviews or compiling information on peaceful anti-war organizations and other social movement actors. There has also been a notable increase in the number of states adopting ecoterrorism laws in reaction to high profile acts of vandalism, raids on research facilities, and other militant activities carried out by groups such as the Earth Liberation Front and Animal Liberation Front. The militant tactics of such movements have resulted in millions of dollars of damage giving some justification for ecoterrorism legislation, however it generally encompasses broad swaths of the environmental and animal rights movements, even groups that denounce violence and property damage (Lovitz, 2007).

Concurrent to tentative legislative coupling of social movement mobilization domestic terrorism was an overall decline in public support for civil liberties following the 9/11 attacks (Sullivan and Hendriks, 2009). Dissent was seen as a challenge to
the effort against terrorism and treated with suspicion (Davis, 2007). As a result, political support for widespread surveillance and intelligence gathering programs on social movement mobilization—which may have been politically risky before 9/11—was no longer a significant impediment to their implementation. Social movement actors were also aware of the lack of public support for anything that could be considered contentious mobilization. For instance, a coalition of groups planned a series of demonstrations to protest the World Bank and International fund on September 29, 2001 in Washington, D.C. After the terrorist strike several key organizers (particularly labor unions) withdrew from the coalition (Gillham and Edwards, 2003). A demonstration supporting a peaceful response to the attacks ultimately occurred, though with drastically fewer numbers than had been anticipated for the original September 29 events.

4.2.3 Summary and Expectations

Covert repression is a quintessential but understudied element of state repression. Previous research on the topic has primarily focused on repression occurring during the McCarthy era and the 1960s wave of protest. Since most previous research has examined patterns of covert repression direct at a small number of social movements, I have proposed two key mechanisms to help explain how covert repression functions more generally: (1) the demands leveled at the organizational tasked with covert repression; and (2) the logic of target selection. Drawing on the contrasts between covert repression during the COINTELPRO era of the 1960s and 1970s and covert repression following the 9/11 terrorist attacks, I argued that contemporary covert repression is grounded in a logic of information gathering unlike the intended neutralization of movements in the COINTELPRO era. This is not
to say that government actors are now willing to ignore actionable intelligence, but instead to hypothesize that state actors are less likely to have direct covert interventions with social movement groups or actors.

The post-9/11 era provided a perfect storm for an expansion of covert repressive programs. The 9/11 attacks resulted in a centralization of key government and law enforcement agencies with an emphasis on intelligence gathering and sharing, at least in principle. Legislation such as the USA PATRIOT Act expanded the definition of domestic terrorism in a way that considerably overlapped with social movement mobilization while public support of civil liberties declined. We should therefore expect that covert repression is relatively common and occurs across a wide variety of groups. Another consequence of the 9/11 attacks was a loss of political support for First Amendment rights coupled with a decline in public support for civil liberties. In contrast to the argument that social movement mobilization has become a legitimate part of the political process (Meyer and Tarrow, 1998), I propose that social movement mobilization itself became politically marginalized following 9/11. This is particularly true for progressive causes as well as movements against the War on Terror or favoring animal rights and the environment.

4.3 Data and Methods

To examine patterns of covert surveillance I use a database of Philadelphia social movement organizations active between January, 1996 and October, 2009 at risk of being surveilled by the Pennsylvania Office of Homeland Security (PA-OHs) between late 2009 and 2010. The political context for the surveillance program is important, since Ed Rendell was Governor of Pennsylvania. Rendell is a strong Democrat and supporter of civil liberties who was not aware that surveillance
was occurring for several months, and suspended the program once he learned of it (Martin, 2010). If such a program was able to be established during Rendell’s tenure in Pennsylvania, it is not difficult to imagine similar programs were adopted in more conducive political climates.

A list of active organizations was obtained using all public collective action events (e.g. protests, demonstrations, or vigils) reported by the *Philadelphia Inquirer*.\(^2\) Full text searches using several broad search terms were used to generate as comprehensive an enumeration of events as possible. The details of each event were coded to capture a wide variety of factors, including the behaviors of the protestors and the police, as well as the claims articulated by participants. Of particular importance for this analysis was the organizational sponsors of each event (if any). Up to four separate sponsoring SMOs were retained for each event.\(^3\) Overall, this strategy yielded a total of 779 distinct collective action events. Of these, I identified a total of 371 SMOs that staged 491 of the events. Protests without a reported organizational sponsor were dropped from any further analysis.

Standard sources of data on social movement mobilization such as newspapers do not provide systematic information about covert surveillance. As a result I supplemented the coded information by cross-referencing the list of Philadelphia SMOs with another list of SMOs that were surveilled by the PA-OHS between between November, 2009 and September, 2010. This latter list was publicly released by the

---

2 Another possibility would have been to use the *Philadelphia Daily News* rather than the *Philadelphia Inquirer*. Comparisons of the depth and breadth of coverage of the two newspapers suggested that the *Philadelphia Inquirer* provided consistently better coverage of protest events. While research by Hocke (1998) suggests that combining newspapers can increase the number of events by up to 25%, only 5% of the events covered by the *Philadelphia Daily News* were not covered by the *Philadelphia Inquirer*. Based on these considerations, I suggest that the use of a single newspaper is defensible in this case.

3 This data is drawn from a larger study of social movement mobilization in 20 U.S. cities between 1996 and 2006. This analysis extends the Philadelphia subsample forward by three years in order to more closely match the surveillance period. A description of the larger sample is provided in Chapter 2.
Pennsylvania chapter of the American Civil Liberties Union (PA-ACLU). The list provided by the PA-ACLU contained reports on surveillance of SMOs across Pennsylvania and I removed all SMOs that were not active in Philadelphia during the analytic period. To do so I relied on a series of newspaper and Internet searches for each organization to determine whether there was any evidence of activity in Philadelphia. While the list of surveilled organizations alone is limited in that it only provides any information on which SMOs were surveilled, when combined with the list of SMOs active in Philadelphia between 1996 and 2009, it is possible to more systematically assess the organizational and behavioral features predicting surveillance by the PA-OHS across a wide variety of groups. The final analytic sample contained a total of 431 distinct SMOs.

Before turning to a discussion of the variables, two issues related to the sampling design I use require additional commentary. First, it is well established that patterns in the type of protest events covered by the mass media are but a fraction of the population of protests. Indeed there is a distinct patterning to the type of events that are represented in newspapers, which McCarthy et al. (1996) refer to as selection bias. Due to the widespread use of newspapers to create databases of social movement activity, several studies have examined the factors affecting whether an event is more or less likely to be covered by the mass media (for the most recent review see Earl et al., 2004). Such research consistently suggests that larger events or those having sensationalistic elements (e.g. arrests or other forms of police force, protestor initiated violence) are much more likely to be covered (McCarthy et al., 1996; Barranco and Wisler, 1999; Oliver and Myers, 1999; Oliver and Maney, 2000; Ortiz et al., 2005; McCarthy et al., 2008). Addi-

---

4 A list of the organizations that were surveilled is available on the World Wide Web. See: www.aclupa.org/downloads/trackedlist11910.pdf.
tional research points toward a regional bias in newspaper data, such that events occurring in close proximity to the editorial offices of a newspaper are also more likely to be covered (Myers and Caniglia, 2004; Ortiz et al., 2005). While such findings raise some serious concerns regarding the use of newspaper data when addressing certain research questions, I suggest that its use is defensible in this instance. By using the Philadelphia Inquirer and focusing only on SMOs active in Philadelphia, the geographic component of selection biased is minimized. As well, due to the widespread institutionalization of social movements in the United States (McCarthy and McPhail, 1998), many SMOs have developed cooperative relationships with mass media actors and are quite sensitive to the factors favored by journalists. This may result in a bias towards events initiated by a formal SMO rather than more spontaneous and likely smaller protest gatherings.

A second concern is the inclusion of an SMO that is no longer active during the surveillance program, which may bias the statistical results. It is possible, for example, that analyzing SMOs that are no longer active may understate the breadth of the surveillance program used by the PA-OHS. To minimize these concerns, several steps were made to cull SMOs that appeared to be defunct prior to the onset of the surveillance period I examine (November, 2010). First, organizations that mobilized for a sole purpose around a unique event (e.g. the RNC meetings in 2000) were omitted. Second, I conducted an Internet search for each organization in the sample and excluded any SMO appearing to be defunct. Third and finally, while conducting the Internet search, SMOs that changed their name or merged with other SMOs over the analytic period were also noted, and only the most current available name was cross-referenced with list provided by the PA-ACLU. With those caveats in mind, I now turn to a discussion of the measures I use in the statistical analysis. I return to the limitations of this sample in the concluding
section.

4.3.1 Variables and Measurement

The response variable used in the analysis below is a dichotomous indicator measuring whether (= 1) or not (= 0) a given SMO was surveilled by the PA-OHS. While it would be preferable to have more detailed information regarding the depth, frequency, and overall degree of surveillance for each SMO, the listing provided by the PA-ACLU was quite limited allowing only a binary operationalization. Although they are quite important, cases of highly disruptive or subversive practices by state actors are very rare, while covert surveillance is much more common. Even though the operationalization used here is limited, it is advantageous in its ability to be applied to a wide variety of groups. As Marx (1974) noted in his seminal study on police infiltration of social movements, “…the nature of the topic requires a greater degree of reliance on unusual data sources and perhaps as a result more tentative conclusions.” (p. 404)

To examine patterns of SMO surveillance, I use a total of three conceptual groupings of predictor variables. The first measures each SMO’s mobilization history, which is intended to collectively account for each group’s level of activity over the analytic period as well their willingness to engage in contentious tactics that may disrupt public order. I begin with a variable measuring the number of protest events stages by each SMO between 1996 and 2009. Next I use a dichotomous variable indicating whether each SMO staged a contentious protest between 1996 and 2009, which includes using any form of civil disobedience, spatial contention (e.g. protesting inside a store), property damage, or violence (1 = yes; 0 = no). Last, I use another dichotomous variable indicating whether the police used arrests
or other forms of force at any event staged by each SMO between 1996 and 2009 (1 = yes; 0 = no).

The second group of variables differentiates between the type of social movement, based on the group's primary issue or topic of mobilization.\(^5\) I use a polytomous measure with eight different movements that is intended to achieve a balance between the breadth of topics around which social movements mobilize and the empirical frequencies in the sample. Since the sample size is relatively small, there were not enough cases to create certain groups that would ideally be included in the analysis (e.g. less than 10 Muslim SMOs were identified). In the end, I differentiate between the environmental/animal rights, conservative/right wing, peace, religious, and anti-violence movements as well as labor organizations. I also use a category for progressive SMOs, which combines groups emphasizing issues including pro-choice, anti-racist, anti-police brutality, or the death penalty. Another category is used for groups mobilizing around identity issues (e.g. race, ethnicity, LGBT, women’s rights). Finally, I use a category for all other social movements that includes SMOs focusing on issues such as education, motorcycling, the Internet, or support for public libraries. This latter group is used as the reference category in all analyses.

The final set of variables measures SMO visibility and uses two covariates. I began with a dichotomous variable indicating whether (= 1) or not (= 0) the SMO had an active website. Such websites generally provided information on previous and subsequent campaigns for each SMO, and thus could provide an

\(^5\) A small number of groups could legitimately be placed into two or more categories (e.g. Muslims Against the War in Iraq). In such cases, the group was coded to reflect the category deemed to be primary substantive importance based on their website (if any), media coverage, and the claims articulated at any protest event. Additional analyses (available upon request) included a dichotomous variable for multi-issue SMOs. This variable did not substantively or statistically improve model fit suggesting the results are robust to such issues.
easily accessible source of information for the PA-OHS. I also use a continuous measure of the number of times each SMO was mentioned in the *Philadelphia Inquirer* between 1996 and 2009 based on full-text searches for each SMO’s name. Due to a small number of SMOs mentioned in several thousand newspaper articles (primarily labor unions), I take the natural logarithm to minimize such outliers.\footnote{Since the natural logarithm of zero is not defined, a constant of 0.1 was added to all SMOs that were not covered in the media.}

### 4.3.2 Analytic Strategy

I use Bayesian semiparametric logistic regression to examine patterns of SMO surveillance.\footnote{There are two continuous variables in the model (number of protests and number of newspaper articles), making it possible that either or both had a non-linear relationship to SMO surveillance. To account for this possibility I used an iterative process where each continuous measure was entered as a non-parametric term along with the other covariates, and then fit a model with a smooth for both variables. Visual inspection of the results suggested that effect of the number of articles is approximately linear, and therefore I only use a smooth for the logged number of newspaper articles.} I begin by assuming that SMO surveillance follows a Bernoulli distribution, such that

\[
\text{SMO Surveillance}_i \sim \text{BR} \left( p_i \right). \tag{4.1}
\]

To convert the probability model in Equation (4.1) to an additive model, I use the logit link function so that

\[
g(p_i)^{-1} = \ln \left( \frac{p_i}{1-p_i} \right) = \eta_i.
\]

This results in the semiparametric model

\[
\eta = X\beta + f \left( \# \text{ of Articles} \right), \tag{4.2}
\]

where \(X\) is a matrix of observed covariates, \(\beta\) is a column vector of parametric effects, and \(f \left( \# \text{ of Articles} \right)\) is a nonparametric smooth.

Following research by Ruppert et al. (2003) and Crainiceanu et al. (2005), I
use a thin-plate smoothing spline to estimate the nonparametric smooth in Equation (4.2). While a thorough technical discussion is beyond my current scope, such procedures have several advantageous properties for Bayesian analysis. As Crainiceanu et al. (2005) demonstrate, Equation (4.2) can be rewritten as a mixed model taking the form:

\[ \eta = X\beta + Zb. \]  

(4.3)

In Equation (4.3), \( b \) refers to \( k \) random effect parameters that are pre-multiplied by \( Z \), an \( n \times k \) matrix of knots with elements

\[ Z = \begin{bmatrix} |x_i - \kappa_1|^3 & \ldots & |x_i - \kappa_k|^3 \\ \vdots & \ddots & \vdots \\ |x_n - \kappa_1|^3 & \ldots & |x_n - \kappa_k|^3 \end{bmatrix}. \]  

(4.4)

In Equation (4.4), \( \kappa_i \) refers to evenly spaced quantile-based knot points over the distribution of the logged number of newspaper articles. A total of 20 knots are used, which are known to provide good approximations of potential non-linearities without overfitting (Ruppert et al., 2003).

Finally, since I use Bayesian estimation, prior distributions are assigned to model parameters. I use the following uninformative priors:

\[ \beta_i \sim \mathcal{N}(0, 0.001) \]
\[ b_i \sim \mathcal{N}(0, 0.001) . \]

To be consistent with the Bayesian literature precisions rather than variances are used in the prior specification above. This translates to assigning normally distributed priors with an expectation of zero and standard deviation of 1000. I use Markov chain Monte Carlo (MCMC) techniques to simulate the joint posterior
density of the model in Equation (4.3). A recent overview of MCMC methods is Brooks et al. (2010) and Jackman (2009) or Gill (2008) provide more general treatments of Bayesian methodology. I use three parallel chains for the regression analysis, each with 50,000 MCMC draws.\(^8\) I estimate the models using the \texttt{rjags} (Plummer, 2011) package in R (Version 2.13.0; R Development Core Team, 2011) to compile and run the models using JAGS (Version 2.2.0; Plummer, 2010).

### 4.4 Results

Table 4.1 provides descriptive statistics for the variables used in the analysis. On average, 22.27\% of organizations in the sample were surveilled by the PA-OHS. This supports the argument that covert repression is currently quite expansive and covers a wide variety of groups. If such patterns generalize across the United States, it is possible that tens of thousands of social movement organizations are being or have been surveilled. Of course a baseline rate of the prevalence of covert surveillance is needed to definitively conclude whether the PA-OHS is more or less aggressive in its surveillance program, although data limitations make such information difficult to estimate. It is also noteworthy that over 21\% of SMOs with a contentious event between 1996 and 2009 were surveilled, which is less than 1\% below the average level of surveillance. Similarly, 24.49\% of SMOs who organized an event where the police used force or arrests were surveilled, which is only marginally larger than the mean level of surveillance. When taken together it appears as though neither a contentious history nor a history of conflictual interactions with the police are a driving factor when law enforcement agencies

---

\(^8\) Since the convergence of MCMC algorithms cannot be conclusively determined in applied contexts, the use of multiple chains allows me to estimate variety of convergence diagnostics (a review of such issues is provided by Jackman, 2009). In all models, there appears to be satisfactorily convergence.
decide which groups to surveil.

Table 4.1. Descriptive statistics for variables used in the analysis \((n = 431)\)

<table>
<thead>
<tr>
<th></th>
<th>Std. Mean</th>
<th>Std. Dev.</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covert Surveillance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMO Surveilled</td>
<td>0.22</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Mobilization History</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Protests, 1996-2009</td>
<td>1.68</td>
<td>2.61</td>
<td>0.00</td>
<td>25.00</td>
</tr>
<tr>
<td>Previous Contentious Protest</td>
<td>0.17</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Previous Police Use of Force/Arrests</td>
<td>0.11</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Type of Social Movement</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Environment/Animal Rights</td>
<td>0.06</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Conservative/Right Wing</td>
<td>0.06</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Peace</td>
<td>0.09</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Progressive</td>
<td>0.18</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Religious</td>
<td>0.05</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Labor</td>
<td>0.12</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Identity</td>
<td>0.17</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Anti-Violence</td>
<td>0.10</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>SMO Visibility</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMO Has Active Website</td>
<td>0.76</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Number of Newspaper Articles (ln)</td>
<td>2.27</td>
<td>2.28</td>
<td>-2.30</td>
<td>7.88</td>
</tr>
</tbody>
</table>

The distribution of the types of social movements points to a relatively high range of diversity in the sample. Progressive movements are by far the most common representing approximately 18% of cases, which is followed by identity movements at 17%. Religious movements are comparatively sparse at 5%. These univariate statistics mask considerable variability in the prevalence of surveillance across the different movements. Figure 4.1 plots the magnitude of the deviations from the overall means for each movement. The x-axis of the figure is centered on the overall average of 22.27%, with below average rates of surveillance taking negative values and conversely above average rates of surveillance taking positive values. For example only 4.86\% (= 22.27 – 17.40) of anti-violence SMOs were surveilled while 57.14\% of environmental or animal rights SMOs were surveilled.
Examining Figure 4.1, it is clear that progressive movements are overrepresented among the movements with above-average surveillance. An exception to this is rightist movements, which are approximately 7% more likely to be surveilled. Religious movements as well as labor, identity, anti-violence, and all other movements are less likely to be surveilled. While all of the movements had some degree of surveillance, it was by far the most common for the animal rights and environmental SMOs, which is consistent with my expectations. Given the lack of effect for previous contentious protests or events with police use of force or arrests, the relatively widespread surveillance of progressive and conservative movements is striking and suggests that contemporary patterns of surveillance are more heavily influenced by protestors themselves rather than their actions.
Table 4.2. Posterior means and standard deviations from a Bayesian semiparametric logistic regression model predicting the surveillance of Philadelphia Social Movement Organizations (n = 431).

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mobilization History</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Protests, 1996-2009</td>
<td>-0.31*</td>
<td>-0.23*</td>
<td>-0.24*</td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
<td>(0.10)</td>
<td>(0.09)</td>
</tr>
<tr>
<td>Previous Contentious Protest</td>
<td>0.29</td>
<td>0.55</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td>(0.42)</td>
<td>(0.44)</td>
<td>(0.46)</td>
</tr>
<tr>
<td>Previous Police Use of Force/Arrests</td>
<td>0.36</td>
<td>0.10</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>(0.46)</td>
<td>(0.48)</td>
<td>(0.50)</td>
</tr>
<tr>
<td><strong>Type of Social Movement</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Environment/Animal Rights</td>
<td>2.65*</td>
<td>2.39*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.60)</td>
<td>(0.66)</td>
<td></td>
</tr>
<tr>
<td>Conservative/Right Wing</td>
<td>1.71*</td>
<td>1.17</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.61)</td>
<td>(0.66)</td>
<td></td>
</tr>
<tr>
<td>Peace</td>
<td>1.71*</td>
<td>1.54*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.56)</td>
<td>(0.61)</td>
<td></td>
</tr>
<tr>
<td>Progressive</td>
<td>1.54*</td>
<td>1.21*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.52)</td>
<td>(0.56)</td>
<td></td>
</tr>
<tr>
<td>Religious</td>
<td>1.19</td>
<td>0.97</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.70)</td>
<td>(0.76)</td>
<td></td>
</tr>
<tr>
<td>Labor</td>
<td>0.97</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.59)</td>
<td>(0.63)</td>
<td></td>
</tr>
<tr>
<td>Identity</td>
<td>0.70</td>
<td>0.51</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.55)</td>
<td>(0.60)</td>
<td></td>
</tr>
<tr>
<td>Anti-Violence</td>
<td>-0.73</td>
<td>-0.65</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.95)</td>
<td>(0.99)</td>
<td></td>
</tr>
<tr>
<td><strong>SMO Visibility</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMO Has Active Website</td>
<td>1.74*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.49)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Newspaper Articles (ln)</td>
<td>See Fig. 4.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.93*</td>
<td>-2.23*</td>
<td>-4.65*</td>
</tr>
<tr>
<td></td>
<td>(0.17)</td>
<td>(0.46)</td>
<td>(1.18)</td>
</tr>
<tr>
<td>DIC</td>
<td>454.20</td>
<td>432.50</td>
<td>388.40</td>
</tr>
</tbody>
</table>

Note: Reference category is all other social movements. MCMC standard deviations in parentheses. *=95% credible intervals do not overlap zero.
Table 4.2 provides the posterior means, standard deviations, and 95% credible intervals for a series of nested models. Each model introduces a new conceptual grouping of coefficients and changes in overall model fit can be compared by the Deviance Information Criterion (DIC). The DIC is a measure of fit proposed by Spiegelhalter et al. (2002) that penalizes for overly complex model specifications. The DIC is scaled such that lower values indicate better model fit and differences in DIC of 10 between two models suggests a notable improvement in explanatory power.

Model 1 introduces the group of coefficients measuring each organization’s mobilization history. The posterior mean for number of previous protests is negative and remains so when the other blocks of coefficients are added in Models 2 and 3. By Model 3 the 95% Credible Interval (CI) is between -0.43 and -0.08 suggesting that staging more public protest events decreases the likelihood of surveillance. While this may appear counterintuitive it might be the that the PA-OHS are more prone to surveil groups whose actions are hidden from public view. Engaging in public protest events makes a clear statement of a group’s activities and positions. The effects for previous contentious protest and previous police use of force or arrest have positive posterior means indicating an increase in the likelihood of covert surveillance. However they each have relatively high MCMC standard deviations and in the full model specification the 95% CIs are between -0.27 and 1.54 for contentious protests and -0.59 and 1.37 for police use of force/arrests. It is therefore impossible to conclude a consistent directionality given this data. These estimates are consistent with the small differences in surveillance rates across these variables noted above and also corroborate Davenport’s (2005) analysis of the Republic of New Africa that also pointed toward the limited effect of dissident behavior.

The block of coefficients measuring the type of social movement is added in
Model 2. This decreases the DIC by 21.7 relative to Model 1 indicating that differentiating between the type of SMO improves model fit. There is a uniformly positive 95% CI for the environmental/animal rights, peace, and progressive movements. The magnitude of the effect size is also largest for environmental and animals rights SMOs. Across Models 2 and 3 the posterior distributions for the religious, labor, and identity movements have positive posterior means but also have wide 95% credible intervals that overlap zero. There is a negative posterior mean for anti-violence groups, however here too there is sufficient uncertainty around the estimate that the 95% CI contains both positive and negative values. The 95% CI for conservative and right wing groups is uniformly positive in Model 2 but is not robustly so after the remaining variables are included in Model 3 (95% CI: -0.08;
When taken together these results point towards an asymmetric preference on behalf of the PA-OHS to surveil progressive movements and the animal rights, environmental, and peace movements as well. There may also be a tendency to surveil right wing or conservative groups—the 95% CI for right wing groups is mostly positive—but this relationship is partially explained away once the proper controls are added. It is also noteworthy that even though the behavior of an SMO or their history of conflictual interactions with law enforcement has little effect on the likelihood of covert repression, the ideological stance of an SMO does have an impact.

Figure 4.3. Fitted posterior median probabilities and 95% credible intervals of the covert surveillance of Philadelphia social movement organizations by the Pennsylvania Office of Homeland Security across the type of social movement from a Bayesian semi-parametric logit model ($n = 431$).
visibility. The change in DIC between Models 2 and 3 is 44.1, suggesting an improvement of model fit and that Model 3 provides the best fit overall. It is clear that having a website increases the likelihood of surveillance considerably (95% CI: 0.83; 2.78). The information provided by the PA-ACLU cannot be used to determine whether the PA-OHS used electronic surveillance or other forms of covert repression. However, since electronic surveillance using a spider or scraper to systematically analyze the website materials is simple and cost effective while being nearly untraceable, it may be a preferred tactic to surveil a wide variety of groups while at little expense. To visualize the relationship between the newspaper coverage and the likelihood of surveillance, a plot of the non-parametric smooth from Model 3 is provided in Figure 4.2. The solid line represents the posterior mean while the dashed lines represent the lower and upper boundaries of the 95% credible interval. There is a complex and highly nonlinear relationship between the two variables where groups that receive no media coverage are the most likely to be surveilled and this decreases as for low levels of media attention. This relationship reverses and slowly increases over the range of the variable as newspaper coverage becomes more extensive, though the 95% CI for the effect as a whole remains mostly negative. The boundaries of the 95% CI also gradually widen, particularly for SMOs with very high levels of newspaper coverage. When coupled with the negative effect of the count of previous protests, this finding again suggests that SMOs who generally operate outside the purview of the general public are more likely to be surveilled. As media attention increases the likelihood of covert repression increases as well.

Given the considerable variance in surveillance patterns across the different social movements under investigation, it is useful to visualize the relative differences. To do so Figure 4.3 provides the fitted posterior medians and 95% CIs across each
different movements. The probabilities were calculated separately for each type of movement (with all other movements set to 0) while holding the remaining other variables constant at their mean. These estimates are based on Bayesian, so the 95% CIs can be interpreted as the probability that a SMO of a particular movement will be surveilled given this data, holding all other variables constant. As should be expected given the regression results, the posterior distribution for environmental and animal rights groups is quite large though with high uncertainty. The 95% CI ranges between 0.33 and 0.77 with a median of 0.56. The effect sizes for peace and progressive movements are somewhat smaller, but still point towards above average levels of covert surveillance. It is also important to note that even though the 95% CI for conservative and right wing groups overlaps zero in Model 3, there is noticeably more uncertainty around the estimate. At the opposite end it is clear that once uncertainty around the estimates is accounted for, the range of probabilities that anti-violence groups are surveilled is quite small with a lower bound of 0.008. The posterior medians for labor and religious movements are roughly comparable in size to that of conservative and right wing groups, though again the range of uncertainty is rather large.

4.5 Discussion and Conclusion

This study set out to examine current patterns of covert repression. I began by suggesting that even though covert and overt repression are part of a single continuum, they have different explanatory mechanisms. Drawing on previous research, I argued that covert repression is influenced by the organizational demands of the repressive actor coupled with a logic of target selection favoring the repression of politically marginalized social movements. Unlike the counterintelligence
programs used during the COINTELPRO era that were intended to disrupt or neutralize social movement mobilization, I proposed that contemporary covert repression evolved towards an emphasis on intelligence gathering instead of direct engagement. This is facilitated by the centralization of major federal agencies with a stated purpose to collect and analyze intelligence coupled with a legislative environment giving wide discretionary power to law enforcement while linking mobilization to domestic terrorism. To examine the support for these arguments I analyzed a database of 431 social movement organizations active in Philadelphia between January, 1996 and October, 2009 at risk of surveillance. The time period I examine is particularly salient since Ed Rendell—a lifelong Democrat with a reputation for respecting civil liberties—was Governor. The results suggest that covert repression is quite widespread, and over 22% Philadelphia SMOs in the sample were surveilled. Covert repression is also directed at the progressive, peace, environmental, and animal rights movements, and groups that operate largely outside of public view. These results have strong implications for developing a cogent theory of state repression and the intersection of social movement mobilization and civil liberties more generally.

A broader question revolves around the symbiotic relationship between covert and overt repression. This research does provide benchmarks regarding the organizational factors that influence the likelihood of covert repression, but cannot speak to its consequences. Therefore a significant issue not addressed in this study is whether the PA-OHS or other law enforcement agencies used the intelligence to interfere with the articulation of First Amendment rights. After conducting an ethnography of 71 social justice organizations Starr et al. (2008) conclude that contemporary covert repression follows a logic and practice similar to the COINTELPRO era. This is inconsistent with my findings, which failed to find any
systematic relationship between the police use of force or arrests as predictors of covert repression. It should be noted that my findings do not preclude the possibility that government actors still rely on counterintelligence tactics to repress social movement organizations and actors. Future research should combine both qualitative and quantitative methodology to further disentangle this complex relationship. It should be emphasized however that even a strict adherence to covert intelligence gathering can have a considerable chilling effect on freedoms of association and assembly. Intelligence gathering is often used for the purpose of extra-legal intimidation of movement actors (Marx, 1974) and can also have a significant emotional toll on individual activists (Cunningham and Noakes, 2008).

It is also important to stress the lack of support for a relationship between a group’s previous use of contentious protest tactics or conflictual exchanges with the police and covert surveillance. The logic of target selection then, at least in this sample, was not based on behavior, but on other factors. This finding contrasts with the well established relationship between threatening behavior during a protest and more coercive police responses noted by several scholars (Davenport, 2007a; Soule and Davenport, 2009). Such findings raise the question of whether the PA-OHS was overly aggressive in its target selection, focusing on groups that perceived to have the capacity to disrupt public safety regardless of their previous record. These findings are also consistent with work by Cunningham (2003a, 2004) suggesting that during the COINTELPRO-era, the decision to repress was heavily contingent on the government actor’s perception of challenging groups, rather than their activities.

A limitation of this research is how my analytic sample may generalize beyond Philadelphia SMOs. While a nationally representative list of SMOs containing information on whether they were surveilled would be ideal, to my knowledge a
national listing of groups that were surveilled does not exist. A degree of caution is therefore required before definitively drawing any conclusions at the national level. Nonetheless, as I noted above, if a covert surveillance program was able to operate for several months despite the political opposition of a state’s leadership, similar programs are likely common across the United States. Indeed, comparing the types of social movement organizations that were surveilled is consistent with national (but unsystematic) trends reported by the Union (2011). Based in this, it is likely the surveillance practices for Philadelphia SMOs is quite comparable to those in other locations. The pattern of results also point towards important questions that may be more adequately tested with with different sources of data. First, the regression analysis suggests an above average level of surveillance for progressive social movements, fitting with concerns raised about the implications of the USA PATRIOT Act. After the Iraq war of 2003, several anecdotal reports across the United States arose with allegations of particularly high levels of covert repression (e.g. Fisher, 2004; Starr et al., 2008). These results provide initial corroboration for such claims as peace organizations were specifically targeted for covert surveillance. Second, certain model specifications also indicated that conservative and right wing organizations were targets of covert repression. While in the case of Philadelphia these effects were not robust, it is entirely plausible that conservative groups are also overrepresented among groups that are surveilled. Given the recent widespread mobilization of the Tea Party it is important to examine whether such observations are true at the national level.

Another limitation is that my sample of organizations contains only skeletal information about each SMO. This makes it difficult to create a comprehensive picture of larger set of the organizational characteristics influencing patterns of surveillance. It is quite possible that factors such as membership counts, income,
office space, radicalness, and other factors may influence the decision of government actors to covertly surveil an SMO. Since research by Cunningham (2003a, 2004) suggests that decisions to repress are heavily contingent on the repressive actor’s perception of challenging groups, the presence or absence of certain resources or ideological positions not captured here may add further explanatory power to the models. More extensive information on the organizational characteristics of groups in the sample would help to more systematically address this question. With that said, the measures used in the analysis allow for two tentative conclusions. Specifically, SMOs that receive little attention from the mass media are more likely to be surveilled, though there is also a penalty for groups with extensive media exposure. It follows that groups with the organizational capacity to produce the right amount and tone of media coverage may insulate themselves from covert repression. A second point is that groups with the organizational capacity to host large numbers of public events are comparatively privileged and less likely to be surveilled. There may be an advantage to more professionalized SMOs, who are less likely to become targets of covert repression due to a more effective capacity to spur mobilization.
Chapter 5

Conclusion

5.1 Synthesizing and Looking Forward

At the beginning of this dissertation I argued that protest requires accessible social space. This has been almost universally recognized by governments in western democracies and often codified in each country’s foundational documents. The core arguments and findings developed above strongly suggest not only that the accessibility of space is contingent on the identity and spatial location of the challenging group, but perhaps that tolerance for protest and dissent has declined significantly more generally. To examine the reasons why there appears to be movement away from the comparatively tolerant, negotiation-based model of protest control (widely expected in the existing literature), I have examined several ways that the police can control dissent. The common theme throughout this project has been to link broader structural, social, and institutional transformations in crime control, policing, and municipal politics to protest policing. At the time of writing, none of these inter-connectivities have been systematically examined, and much protest policing theory has focused almost exclusively on social movement dynamics alone.
This has created a movement centric bias in protest policing and my dissertation poses a significant challenge to this narrow conceptualization. Existing explanations of protest policing also have a rationalist bias. Protest policing is generally assumed to reduce to law enforcement agencies using an amount of force that is roughly proportionate to the level of threat posed by the demonstrators (Davenport, 2007a). This research also begins a more systematic critique of the rationalist explanations.

Chapter 2 examined the policing of protest in 20 major cities between 1996 and 2006. I linked the extensive variation in protest policing across the cities to structural sources, particularly the differential adoption of broken windows policing strategies, training and technological innovation in policing, and the political context of each city. Chapter 3 was more explicitly temporal and examined protest policing in New York City between 1960 and 2006. I developed the concept of policy spillover to explain how municipal policies and institutional changes unrelated to protest still contributed to more aggressive protest control. Chapter 4 examines the covert surveillance of Philadelphia social movement organizations between 2009 and 2010. I demonstrated that covert surveillance has little to do with a group’s contentious history, but is strongly shaped by their ideology.

While I have shown that the policing of protest is shaped and directed by its social context, this merely represents the tip of the iceberg. This dissertation is intended to provide the start of a broader conversation engaging with other streams of research—particularly in social stratification, human geography, the sociology of law, and policing—that have much to contribute to how protest control is understood. In short, it is time for scholars to pay more attention to the intersectionalities inherent in protest policing and move beyond building theoretical models that draw solely from the realm of social movement theory. I conclude
this dissertation by first outlining three major implications of this research that contribute to existing theories of protest policing. I then outline several steps for future research that will extend, refine, and improve what I have begun in this project, and finally close with an argument that disparities in protest policing is a form of political stratification and political inequality in First Amendment rights.

5.2 Major Implications

When analyzed collectively the central results from each substantive chapter have several implications for existing scholarship on the policing of protest as well as for future research. The integration of a broader range of scholarship into protest policing, as I have accomplished here, is far from a simple task so I mention its limitations and/or complications where appropriate. I begin by outlining three major implications and limitations of this dissertation.

5.2.1 Implication 1: Protest Policing Isn’t Always Reactive

One of the core findings that spans this project is a serious challenge to the theoretical orthodoxy that protest policing is mainly a reaction to disruptive or violent behavior. Even though the invocation of situational threats to explain police behavior has become nearly axiomatic, granting such factors theoretical primacy also involves making implicit assumptions that create an image of protest policing inconsistent with large swaths of the literature on law enforcement operations. Focusing on what the protestors do is contingent on the assumption that the protest policing is primarily reactive, despite voluminous examples of racial profiling (e.g.
or policy decisions asymmetrically affecting the poor or socially marginalized (Choongh, 1998; Wacquant, 2009). It is unclear why one should expect the police to stringently adhere to reactive strategies during protest events, but adopt more aggressive proactive strategies elsewhere.

Given the strong theoretical and empirical linkages between a variety of non-behavioral elements of a protest and police responses, this dissertation provides conclusive evidence that such explanations are, at best, overly simplified and incomplete and, at worst, reductionist and misleading. This is not to say that what protestors do during an event is irrelevant; clearly this is not the case. However, my point is that acknowledging police discretion, including both proactive and reactive strategies, are necessary to construct a comprehensive explanation of protest control.

5.2.2 Implication 2: Context Matters

A second major implication of this dissertation is that protest policing is largely a product of its social context. As we saw in Chapter 2, there are extensive differences in how protest is managed across the United States, indicating that there are largely unrecognized inequalities in First Amendment rights. Despite the widespread decentralization of policing common in many western democracies, it is reasonable to expect that otherwise comparable demonstrations occurring in different locations should be handled in a consistent way. This is not the case: even after introducing a wide variety of structural variables, extensive variability remains. Chapter 3 makes it clear that the temporal context is also important. This has been recognized by other scholars (e.g. McPhail et al., 1998; Soule and Davenport, 2009), but prior research has focused almost exclusively on factors directly
influencing protest policing. By looking more broadly at municipal regulations and policies that may appear to have little direct relevance to protest policing, I was able to develop a rationale for why we should expect an increase in police severity over time that was supported by the data. Chapter 4 also makes use of context, particularly the development of the Department of Homeland Security after the 9/11 terrorist attacks, which significantly broadened the sorts of social movement organizations deemed necessary to surveil.

Even though there is little question that protest policing theory would benefit from a more contextualized understanding of how structural conditions and institutional arrangements influence police responses, great care is needed. Paying attention to structural phenomena can quickly devolve into throwing a large number of variables in a model until $p < 0.05$. Much like the use of political opportunities in social movement scholarship, proper theoretical attention to structure in protest policing must avoid launching another “all encompassing fudge factor” (to borrow from Gamson and Meyer, 1996). The most straightforward way to avoid making structural analyses of protest policing so expansive that they become of little use may be to begin with detailed historical analyses or ethnographies of a small number of cases to identify potential structural factors, then use large sample quantitative studies to assess how they scale upwards.

### 5.2.3 Implication 3: Ideology and Identity Matter

A final major implication from this research has to do with the effects of protestor ideology and identity—whether it related to political, racial, ethnic, sexual orientation or any other social category/identity—on protest control. Perhaps the strongest case for the importance of identity and ideology is the results in Chap-
ter 4, which demonstrated considerable differences in the types of social movements that were surveilled by the Pennsylvania Office of Homeland Security. Chapter 3 also suggests that minority rights or peace protests are policed more heavily. Taken together, there is evidence that the police do not control events with an even hand. Due to the over-emphasis on behavior in current social movement scholarship, such considerations are often treated as theoretically marginal or ignored entirely, but it is important to emphasize that the foundational documents of nearly all western democratic countries—such as the U.S. constitution or Canadian Charter of Rights and Freedoms—expressly prohibit discrimination on ideological grounds. There is a disjuncture between de jure and de facto understandings of political expression, and we should take serious notice of it.

The ideology of the challenging group or their identity can also have more subtle influences on protest policing, particularly when police decisions are in a legal gray area. When has a sit-in lasted for “too long?” How closely should the parameters of a protest permit be adhered to? Should civil disobedience be tolerated? In cases where a range of legitimate decisions are subject to police discretion, the political and social identities of the protestors may have a decisive role in police conduct. The pattern of decisions made by law enforcement are central distinguishing elements between despotic and democratic states. It follows that systematic differences in police conduct grounded in the characteristics of the protestors bodes situates a nation-state closer to the former than the latter.

5.3 Future Directions

In research intended to coherently conceptualize protest policing and state repression more broadly is research, Earl (2003) suggested an orienting perspective that
differentiates between the identity of the repressive actor, the character of the repressive action, and whether the repression is overt or covert. Yet it is clear from the discussion above that these while these foci are useful, they are not sufficient to capture several major components of protest policing. The current research is intended to push existing theories of protest control in new directions by drawing from other literatures. Doing so also opens up several avenues for future research to improve existing scholarship, as well as solve puzzles pertinent even within current explanations of protest control. Below I discuss several directions that may be fruitful for future research on protest policing.

5.3.1 Theorizing Space

A lurking concept throughout this dissertation is the importance of space, not only the social space needed for protest to occur, but also how expectations of the spatial routines appropriate for a particular location influence police behavior. Few scholars have attempted to explicitly incorporate the insights from critical geography and social theory into scholarship on social movement dynamics, making this an excellent place for future research to use insights from these perspectives. The most comprehensive attempt to connect theories of space to protest policing is by Starr et al. (2011), who argue that the police have effectively declared war on protestors and act accordingly when managing protest sites. However, they focus strictly on the global justice movement and major protest events, which significantly limits the applicability of their arguments as discussed in Section 1.3.1 above. Additional work by Tilly (2000) on the spatial elements of contentious action has mainly focused on the tactical repertoire used by protestors, again limiting its utility as a general framework. Similar to Starr et al. (2011), I suspect
that Lefebvre’s (1991 [1974]) distinction between representations of space (space constructed through symbols and images), spatial practice (the social reproduction of space), and representational space (the intersection of the first two types) might be particularly useful. Research examining how the spatial distribution of protest, using geocoded samples of events for example, and how this affects police responses might provide a first step to more spatially aware (inductive) theories on protest control.

5.3.2 Expanding the Theoretical Gaze

A core theme in my approach for this dissertation was to use insights from sources outside of social movements scholarship to thicken and extend existing explanations of protest policing. I have focused primarily on research examining policing, social control, and crime control, but several other areas on sociology might provide valuable insights. For example, research on social stratification might point to theoretical pathways through which we can create better explanations of protest control. In Chapter 2 I found that arrests during protests were less likely when the event was held in a wealthier city, suggesting a connection between socio-economic context and police conduct during demonstrations. There is a rich literature suggesting a complex, but inexorable relationship between social inequality and political inequality. Canonical research by Verba et al. (1995), for instance, finds that income is a powerful factor influencing political participation, and this appears to be the case for protest policing also. The gap in political participation will almost certainly widen further in the aftermath of the Supreme Court’s ruling in Citizens United v. Federal Election Commission ruling, which allows unlimited political campaign donations by private businesses through Super Political Action
Committees. The potential implications of this decision cannot be overstated, and could create incentives favoring a patron and client system in Federal politics. Future research should explore and evaluate these connections more systematically than I have done here.

There is also relatively little research on the aftermath of overt police repression (but see Earl, 2005). Arrests, and especially mass arrests, are relatively common components of contemporary demonstrations, and research on incarceration might provide theoretical expectations suitable for research on contentious political activity as well. Illustrative questions that could be examined include: how does incarceration influence activist careers and organizational mortality, in that activists may be less likely to participate after their releases from prison? Are there sentencing disparities for politically motivated criminal convictions relative to other convictions? Are there disproportionately high rates of convictions, reversals, and acquittals when protestors are brought to trial? Scholarly attention to these topics would make our understanding of the social control of protest more holistic and consistent.

5.3.3 Much Ado About Size

The results presented here also provide a significant challenge to previous literature suggesting that larger events are policed more heavily (Davenport, 2000; Earl et al., 2003; Earl and Soule, 2006; Davenport, 2007b; Soule and Davenport, 2009). In Chapter 2 I found a complex, non-monotonic relationship between arrests and event size, where the probability of arrests initially declined as event size grew, before becoming slightly more common once 100 or so participants were present,

---

1 In the case Citizens United v. Federal Election Commission, the Supreme Court of the United States ruled that government restrictions to political funding by corporations, unions, and other organizations violated the First Amendment, and were thus unconstitutional.
though there was considerable variation in police conduct as the crowd size grew. The results in Chapter 3 point to a monotonic but non-linear increase in the likelihood that the police would use force during New York City protests, but event size was unrelated to the arrests. Some of my collaborative research—notably using the same data source as many previous studies suggesting an affirmative relationship between size and force—finds similarly complex effects (see Rafail et al., Forthcoming). To the degree that it is possible to construct valid data on demonstration size, these contradictory findings should be examined much more extensively. The results presented here make it clear that assuming event size remains constantly threatening over time is erroneous, making it important to know when, why, and how event size might influence police conduct.

5.3.4 On Crime Control and Protest Control

This dissertation is among the first scholarly examinations of the effects of crime control on protest control. Yet, there is an apparent inconsistency in the findings suggesting that their relationship is quite complex. Specifically, in Chapter 2 there was a positive effect of property crime clearance rates on the likelihood of arrests, however, in Chapter 3 there was no evidence support a direct relationship between crime rates the use of force or arrests in New York City protests. In both chapters, I initially used non-parametric smooths to ensure that complex non-linear relationships were not washed away by overly strict (and empirically false) parametric assumptions, though the effect of crime clearance rates and crime rates were roughly linear in both cases. The key difference is clearly that Chapter 2 emphasized clearance rates, while Chapter 3 focused on reported crimes.² This

² This discrepancy is due to data limitations, as the clearance statistics were not available for the early 1960s. It was also not possible to create reliable proxy measures by aggregating across
suggests that police priorities, to the degree that clearance rates provide a valid proxy measure, are of more importance than levels of crime. As well, it is possible that relative differences in crime rates across locations are more important than absolute differences in a sole city. Future research should examine and unpack these complexities more systematically. I did consistently show, however, that strategies of crime control such as broken windows can have strong effects on protest policing, but more replication is required to validate these findings.

5.3.5 What About the Police Perspective?

A feature almost wholly absent from this research (and many other studies of protest policing) is direct communication with law enforcement agencies to incorporate their understandings of protest policing into existing theory. While it is true that access to the police requires the careful cultivation of trust, which can be quite time consuming, future research would benefit enormously by engaging with the police more thoroughly. The perception of the police is often assumed with little supporting evidence or discussions with the police officers or managers making the decisions. The blue-centered perspective on protest policing outlined by Earl and Soule (2006), for instance, is predicated on several key assumptions about what the police find as threatening—that are likely correct—but these are not based on anything other than conjecture. The level of access to the police attained by McPhail et al. (1998) or Waddington (1994), is exceedingly rare, but extraordinarily valuable to refining and giving the necessary perspective to how protest policing is understood by those on the front lines and in the command room.

New York State.
5.4 Conclusion

As I conclude this dissertation, I am puzzling over identifying a unifying theme more engaging than “context matters” or “the police are not solely reactive.” While I believe such statements are true, they do not provide a fair assessment of the social and human consequences that heavy-handed state repression can take on political freedoms. As De Tocqueville (2004 [1835]) so cogently recognized, unfettered association life is a vital component of liberty and democracy. Significant social trust is placed in the state and its agents to maintain social order in a democratic society, but there is a parallel expectation that the state will respect those who disagree. This social contract ensures the state retains a legitimate monopoly on the use of force, to borrow from Weber (1991 [1919]). When the social contract is violated, the legitimacy of the state’s use of force is, at best, questionable. I wish to conclude, therefore, by arguing that this dissertation is an examination of political stratification and political inequality, which is much broader than protest policing alone.

The language of social stratification might be a better way to frame inequalities in political access resulting from differential policing decisions made during protest events. Duncan (1968, also see (Tilly, 1999)) suggests distinguishing between inequality and stratification—respectively referring to temporary and durable forms of inequities in power and resources between social groups—and this conceptualization is quite pertinent here. Using Duncan’s terminology, the extensive disparities in protest policing I have identified here are best conceptualized as forms of both political stratification and political inequality. It is stratification in the sense that the First Amendment (and comparable freedoms in other locations) provides a high and consistent bar for police behavior that is routinely violated for reasons
other than public safety. This is inconsistent with the law and spirit of a free and open society, and serves to make certain political acts and speech more protected than others. It is political inequality because one's geographic location, temporal era, or political inclinations strongly condition how and why the police may react. This too is inconsistent with the social contract taken for granted in a democratic society. The movement away from political equality is extensive but potentially reversible if the principles of political equality as core societal values outweigh the state’s interests in limiting dissent.
Appendix A

Statistical Methodology

A.1 Introduction

This appendix serves two central purposes: (1) To provide adequate space to discuss the epistemological and methodological conventions that guide the quantitative components of this project; and (2) To outline the more technical but important details of my analytic strategy in a centralized location in hopes of minimizing digressions into distracting mathematical formulae in the main text.

As is described in more detail below, my primary analytic tool is based on generalized Bayesian semiparametric mixed models. At the time of writing, I have yet to see an application of this class of models in sociology. To justify my decisions, I begin by first developing three general heuristics that I argue are useful in applied social statistics. I then turn to a more extended discussion of the pitfalls that come with using $p$-value based inference and assuming linear statistical relationships as a default. Using that discussion as a springboard, I provide a brief overview of the Bayesian statistical paradigm, and then finally turn to a more detailed overview of my analytic strategy.
Throughout this appendix, I adopt the following stylistic conventions: Matrices and vectors are denoted by bold characters, such as $\mathbf{X}$ or $\mathbf{\beta}$. A superscript $T$ is used for transposition (e.g. $\mathbf{x}^T$). Random variables are denoted in lower-case Roman letters, such as $\mathbf{x}^T = \{x_1, x_2, \ldots, x_n\}$ and random matrices are denoted by uppercase Roman letters. In the discussion of frequentist inference, sample statistics and population parameters are respectively denoted by $\hat{\theta}$ and $\theta$. To reduce unnecessary notation, however, I otherwise drop this distinction in other sections unless a clear differentiation between statistics and parameters is required. When discussing multilevel models in particular, I use $\mathbf{X}$ for matrices of fixed predictor variables and $\mathbf{Z}$ for indicators mapped to random effects parameters (and, as discussed below, for smoothing splines). Similarly, $\mathbf{\beta}$ is used for fixed coefficients and $\mathbf{b}$ is used for random effects. I represent single non-parametric smooths by $f(\cdot)$ and multiple smooths by $f(\cdot)$. As a final caveat note that several mathematical details are intentionally elided, and I generally avoid argumentation by proof or theorem. Since much of what follows has been well established theoretically and in practice, though there are elements of novelty, my intention is provide an informal exposition rather than technical justification.

A.2 Guiding Heuristic Principles

As the positivist strain of quantitative methodology has achieved all but axiomatic status in sociology, it is useful to begin by developing a more reflexive account of the process of quantifying—and thus simplifying—social processes. My purpose here is to outline general heuristics that I use as guiding principles in the statistical modeling. My goal is to engage in non-positivist empiricism without sacrificing rigor. What follows is far from an attempt to build a comprehensive and exhaustive
epistemology, but rather to sketch out some rough-and-ready positions that are widely accessible and reasonably straightforward to implement in practice.

A.2.1 The Box Heuristic

The first heuristic rests on the recognition that mathematical models are, in the best and rarest case, a considerable simplification of more complex (social) processes. As a shorthand, I refer to this as the Box heuristic. The inspiration comes from George Box, who famously stated: “Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful” (Box and Draper, 1987, p.74).\footnote{Even though the reference is drawn from joint work with Norman Draper, the quotation I use is universally attributed to Box alone in the literature. I follow this convention.} While the Box heuristic may seem sophomoric or obvious, even the most casual perusal of empirical sociology will yield uncountable claims to the contrary. Examples include insisting that one has developed the objective measure of an immeasurable social phenomenon, the major statistical corrective that comes with using a model happening to be en vogue at the moment, and the regular use of Bayesian logic to interpret frequentist relationships. Agger (2007) documents dozens of examples of such statements. Similar comments apply to many of the implicit assumptions propping up causal modeling (after all, if a model is definitionally wrong, why should we trust its causal implications?).

There are major implications if the Box heuristic is adopted. Most primarily, if all models are wrong then it follows any given problem has an infinite number of solutions. Clearly some solutions are preferable to others, but it is impossible to definitively conclude that one’s results are sound. When interpreting a statistical model our rhetoric should reflect this fact. Additionally, the Box heuristic implies that even the most promising advancements in applied statistics are imperfect and
contingent on a wide variety of untested (and generally unstated) assumptions. The best model cannot be a substitute for sound social theory. Overall, the Box heuristic serves as a stark reminder that one must retain a healthy skepticism of modeling social phenomena, and instead shift one’s focus to building models that are theoretically driven, open to debate, and falsifiable.

A.2.2 The Berk-Freedman Heuristic

The second major heuristic that I use is referred to as the *Berk-Freedman heuristic*. The central idea stems from Berk and Freedman’s (2003) chapter titled “Statistical Assumptions as Empirical Commitments,” a scathing critique of how social scientists approach the idea of sampling and generalization. Berk and Freedman convincingly argue that assuming a data generation process is random entails significant commitments about the empirical world. When such assumptions are assumed but are untrue—as is often the case in the social sciences—it can result in considerable bias in any statistical analysis. Nearly all statistical modeling in the social sciences proceeds under the assumption that the data was generated by a random process. Yet random samples are almost always the exception, and rarely (if ever) are true random samples available for analysis.

To conduct a genuinely random sample would require, as a matter of course, conducting a census of the population under investigation as a first step. After assigning equal probabilities of selection to the population one must also force all selected respondents to answer all questions accurately. As Berk and Freedman (2003, p. 256) note, collecting a random sample in a country such as the United States would involve the suspension of several constitutionally mandated rights and freedoms. Under its definitional sense then, even major surveys purporting to
represent a random sample objectively fall short of the bar required to defensibly make such statements. The same is even more blatantly true for convenience samples.

Much like the Box heuristic, if one accepts the Berk-Freedman heuristic there are enormous implications. It is primarily a reminder that deficiencies exist in all samples, including everything from pseudo-random surveys to the newspaper data used in this project. Issues of sampling are crucial components of rigorous generalization, and need to be taken more seriously. If one grants the Berk-Freedman heuristic, there are also major implications for statistical inference: if truly random samples are but a theoretical construction, then the theoretical properties of frequentist estimators do not apply in many situations.

### A.2.3 The Ockham Heuristic

The final heuristic is named after the famous medieval scholar William of Ockham. Ockham is perhaps most widely known for creating the philosophic razor bearing his name. Plainly stated, Ockham’s razor is an ontological perspective favoring simplicity over complexity if two explanations have equal explanatory power. In sociology this sentiment has been entrenched as the “law of parsimony,” which is generally used as grounds to eliminate coefficients that are not statistically significant from regression models. As a direct result of such practices, I suggest that parsimony has largely become an end in itself and one can regularly see the omission of theoretically important variables in the name of parsimony if they happen to be insignificant. This tendency violates the spirit of Ockham’s razor and it can lead to considerable consequences in applied work. A lack of statistical significance cannot be conflated with a lack of substantive significance, and omit-
ting explanatory variables is equivalent to making the bold assumption that all omitted variables have no explanatory power. But of course the lack of statistical significance can be due to a variety of factors including a small sample size, undiagnosed violations of model assumptions, measurement error, or a genuine lack of effect. Statistical significance is also incapable of differentiating between an effect that is 0 and the case where 0 falls in the plausible range of values. The power of Ockham’s insight comes from its provision of a clear way to differentiate between two equal explanations. Omitting predictors simply due to a lack of statistical significance or making decisions solely for the sake of parsimony does not come close to meeting this tall standard.

These comments can be justified formally. Let $y$ be a Gaussian-distributed random variable generating the explanatory model

$$ y = X\beta + \omega. \quad (A.1) $$

In Equation (A.1) $\omega$ is an additive composite of a random disturbance term ($\epsilon$), a matrix of omitted predictor variables ($A$), and their corresponding vector of coefficients ($\theta$). The usual assumptions for linear regression apply so that $E(\omega) = \omega^T 1 = 0$ and $E(\text{Cov}(\omega, X)) = 0$. With this new equivalence the regression equation above can be rewritten as

$$ y = X\beta + A\theta + \epsilon. \quad (A.2) $$

It is clear from Equation (A.2) that removing a vector of explanatory variables from the model is substantively identical to assuming that $a^T \theta = a^T 0 = 0 \forall a \in A$. In other words, it assumes that the net effect of all excluded predictors is zero, which
is impossible to conclude using statistical significance as a decisive criteria. The mathematical exposition can also be pushed a step further to demonstrate that omitting statistically insignificant but substantively significant effects might also influence the estimate for the remaining coefficients. Since the analytic solution for linear regression models is

$$\beta = (X^T X)^{-1} X^T y,$$  \hspace{1cm} (A.3)

the omission of one more explanatory factors will influence the calibration of the regression coefficient. The effect of all predictor variables is taken into account since one must solve \((X^T X)^{-1}\).

For my purposes, the Ockham heuristic has two main implications. First, the reading of Ockham’s razor that I propose here implies that statistical significance is not a useful criteria for model building. Instead a much more theoretically and empirically rigorous approach is needed that is grounded in theory and logic rather than arbitrary designations of significance. Second, the methodological tendency to favor parsimony for its own sake must be avoided at all costs. This latter point is particularly relevant to my use of semiparametric smooths, which are often criticized for overfitting or taking up too many degrees of freedom. As I explain below, there are several compelling reasons to avoid assuming that the impact of a continuous variable is best represented by a straight line, even if it requires considerably more degrees of freedom to approximate the relationship.

---

2 Consistent with the Bayesian argument to follow, one could alternatively refer to omitted variables as special cases where completely informative deterministic priors are used that set the marginal posterior density for a particular effect to zero. The end result is exactly the same.
A.2.4 Summary

The three heuristics I outlined above, when taken together, call for a reconsideration of the basic way that social statistics is generally conducted. They are grounded in my own observations that mainstream quantitative sociology is at a crisis point in terms of regular violations of basic statistical theory and the reification of statistical modeling. While I am certain that many would object to various arguments above, at minimum these heuristics serve as the beginning of a badly needed discussion required to produce more defensible quantitative social science. I now turn to a critical review of frequentist inference and assuming linear statistical relationships as the default specification. These sections provide the context for the discussion of Bayesian methodology and my general analytic strategy.

A.3 The Problems with P-Values

The most common approach to statistical inference in sociology is the use of $p$-value based statistical inference generally using thresholds of 0.05, 0.01, and 0.001. The purpose of this section is to advance the position that $p$-values alone provide neither credible or interpretable results for many problems in sociology, and this is particularly the case for the type of data used in this dissertation. The reification of statistical significance through the use of uncalibrated $p$-values—as is currently the norm in sociology—leaves considerable room for improvement. One uses statistical inference to advance certain points, yet when properly understood frequentist inference is rhetorically, technically, and substantively incapable of answering direct questions about the probable range of an effect at the desired level of certainty. Unfortunately this is generally the most basic question that nearly all quantitative research attempts to grapple with.
Before proceeding further it is important to add the following caveat: I am not arguing that all \( p \)-value based inference is necessarily incorrect nor that it should never be applied. There are several strengths to frequentist reasoning and certain frequentist properties are used extensively in the Bayesian approach that I propose below. Instead I believe that a more critical perspective is needed and ignoring the cascade of critical analyses on frequentist inference is becoming increasingly difficult. A thorough literature review would yield thousands of critical citations, but representative examples can be found in a variety of disciplines including economics (McCloskey, 1985; Poirier, 1988; De Long and Lang, 1992; Ziliak and McCloskey, 2008), political science (Gill, 1999, 2008; Jackman, 2009), psychology (Rozeboom, 1960; Cohen, 1990, 1994), sociology (Berk et al., 1995; Raftery, 1995; Lynch, 2007), and statistics (Cox, 1958; Freedman, 1983; Nelder, 1999).

A.3.1 The Logic of Frequentist Inference

I begin with a necessarily incomplete sketch the basic logic of the frequentist paradigm, intended to summarize the central tenants of \( p \)-value based inference. In the textbook case one begins with a \( p \) dimensional vector of statistics \( \hat{\theta} \) which are estimates of the (true) population values \( \theta \) drawn from a simple random sample. The goal is to infer \( \theta \) from \( \hat{\theta} \). It is assumed that \( \hat{\theta} \in \mathbb{R}^p \) and that the elements of \( \hat{\theta} \) are independent and identically distributed.

It is most straightforward to proceed by example, so to begin the exposition let \( \hat{\beta} \) refer to a \( p \)-dimensional vector of linear regression coefficients (though this applies more generally). The data generating process is

\[
y \sim \mathcal{N} (X\beta, \sigma^2 I),
\]
and it follows that
\[ \epsilon = y - X \hat{\beta}, \]
and that
\[ \sigma^2 = \frac{\epsilon^T \epsilon}{N - p - 1}. \]

Equation (A.3) above provides the estimating equation for the regression coefficient and the coefficient covariance matrix is
\[ \Sigma_{\hat{\beta}} = \sigma^2 (X^T X)^{-1}. \quad (A.4) \]

The goal is to infer about the true value \( \beta \) using the values of \( \hat{\beta} \) provided by a single random sample.

The first step in frequentist inference involves linking the individual realization of the true value—\( \hat{\beta} \) in this case—to its parent probability distribution. To do so one can make use of the central limit theorem, which states that for any randomly distributed variable \( \theta \) with expectation \( \mu \) and variance \( \sigma^2 \),
\[ Z_n = \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^{n} \theta_i - \mu \right)/\sigma. \quad (A.5) \]

It follows that
\[ \lim_{n \to \infty} Z_n \sim \mathcal{N}(0, 1), \quad (A.6) \]
meaning that \( Z_n \) asymptotically converges to a standard Gaussian distribution. This holds regardless of the distribution of \( \theta \) and for multivariate distributions as well. Proofs for these arguments can be found in a variety of works (e.g. Wackerly et al., 2008).

From Equation (A.5) we see that \( E(y) = \mu \), and since \( \mu = X \beta \), the expected
value of \( \hat{\beta} \) can be written as

\[
E \left( \hat{\beta} \right) = (X^T X)^{-1} X^T \mu \\
= (X^T X)^{-1} X^T (X \beta) \\
= \beta
\]

Using the central limit theorem, the sampling distribution of \( \hat{\beta} \) is therefore

\[
\hat{\beta} \sim MVN(\beta, \Sigma),
\]

and it is clear that

\[
\begin{bmatrix}
E \left( \hat{\beta}_1 \right) \\
\vdots \\
E \left( \hat{\beta}_k \right)
\end{bmatrix} -
\begin{bmatrix}
\beta_1 \\
\vdots \\
\beta_k
\end{bmatrix} =
\begin{bmatrix}
0 \\
\vdots \\
0
\end{bmatrix}.
\]

(A.7)

The keys to the logic of frequentist inference come from the implications of Equations (A.5) though (A.7). Since we know that the sampling distribution of a set of regression coefficients is asymptotically multivariate normal, we can then assign the probability of drawing a specific \( \hat{\beta}_i \) from its sampling distribution. The value for \( \beta_i \) represents the null hypothesis and is chosen by the analyst (in practice a default of zero is used that is intended to indicate a lack of effect). The final step is then simply of matter of calculating

\[
\int_{-\infty}^{\hat{\beta}_i} \Pr \left( \hat{\beta}_i \right) d\hat{\beta}_i = \frac{1}{\sigma \sqrt{2\pi}} e^{-\left(\hat{\beta}_i - \beta_i\right)^2 / (2\sigma^2)}. 
\]

(A.8)

Equation (A.8) provides the \( p \)-values that are used as the basis for differentiating between significance and insignificance, given then null hypothesis. From this
discussion it follows that $p$-values are but purely theoretical constructs referring to the probability of drawing a random sample with estimates $\hat{\beta}$, given the mean of the sampling distribution (i.e. $\beta$). The sampling distribution remains unobserved, and the value of $\beta_i$ is but one of an infinite number of options.

A.3.2 Criticisms of P-Value Based Inference

A wide variety of criticisms have been raised against frequentist inference, though it is by far the dominant inferential paradigm in applied statistics. Gill (1999), for example, shows that the logic of frequentist null hypothesis testing violates the principle of modus tollens that is heavily used in symbolic logic. A thorough review of such issues is far beyond my scope so I focus my comments on two criticisms of frequentist inference that I suggest strike at the core underlying issues. First, I discuss the utility of frequentist inference when the data generation process is not random or cannot be replicated. Second, I turn to the limitations of interpretability of frequentist inference, resulting in a tendency to treat frequentist probabilities as Bayesian.

A.3.2.1 Frequentist Inference Without Random Sampling

The discussion above makes it clear that frequentist inference requires two components: (1) a random sample generating a statistic of interest; (2) the ability to link the random sample to a purely theoretical normal distribution whose expectation is the population value of the statistic. When either or both of those conditions are not met, $p$-values are quite difficult to interpret or of questionable interpretability.

To justify this argument first consider the scenario where a sample is non-random, but theoretically is replicable (i.e. can produce a sampling distribution).
As I demonstrated above, frequentist inference is only meaningful when $E(\hat{\theta}) - \theta = 0$. This is justified since we can use the central limit theorem which suggests that $\theta$ follows a normal distribution allowing us to integrate over the distribution to assign $p$-values. In cases where $\hat{\theta}$ is not drawn randomly, this no longer holds. Instead the sample statistic can be represented by $\hat{\theta} + \epsilon$, where the latter term contains sampling bias and potentially measurement error as well. Unless one can make a cogent theoretical argument for why $E(\epsilon) = 0$, then there is clearly substantial bias in the $p$-value making frequentist inference untrustworthy.

A second problem emerges when the sampling distribution itself is undefined. This occurs in a variety of contexts such as examinations of nation-states, official crime rates, newspaper coverage of protest events, or more generally to what Berk et al. (1995) refer to as apparent populations. The sampling distribution of $\beta$ is based on a theoretically limitless number of random samples, but the nature of the data drawn from such sources limits the analyst to a single non-random sample.\(^3\) It follows that $p$-values are difficult, if not impossible to interpret. The calculation of a $p$-value involves integrating over the normal distribution—as was noted in Equation (A.8)—which requires the distribution to exist on a theoretical basis. As the name suggests, frequentist inference can only be understood in terms of long-run probabilities over a theoretically justified sampling distribution. In the case of apparent populations this is no longer the case and it is unclear exactly what one is making inference to. As a result, Equation (A.8) cannot be solved and the $p$-value is undefined. Such issues are rarely, if ever, addressed in the existing literature and pose a particularly vexing problem to the regular use of frequentist inference.

\(^3\) While one could be justified in claiming that the apparent population may represent a true population making inference unnecessary, it is exceedingly rare to do so in practice.
A.3.2.2 Bayesian Interpretations of Frequentist Relationships

A longstanding objection to the use of \( p \)-values is their frequent misinterpretation in the literature (Gill, 1999). I demonstrated above that the necessary and inescapable truth of frequentist inference is that \( p \)-values can only be understood in reference to long-run characteristics of the sampling distribution. The mechanical interpretation of a \( p \)-value is: the probability of observing \( \hat{\theta} \) over an infinite number of trials given \( \theta \) is correct is \( p \). The corollary is the frequentist confidence interval, which can be interpreted mechanically as: 95% of confidence intervals around \( \hat{\theta} \) will contain the true value \( \theta \). As I argued above, however, in practice one is nearly always interested in making more direct conclusions: what is the best guess of an effect? What is its plausible range? Frequentist inference, definitionally, cannot be finessed to answer either of these questions even in the vanishingly unlikely scenario that the data is drawn from a fully random sample or the null hypothesis is correct.

Despite the lack of theoretical justification it is routine for analysts to interpret \( p \)-values as the probability that a hypothesis is true or false, and that lower \( p \)-values capture stronger effects (Jackman, 2009; Schrodt, 2010). It is also common to see references to confidence intervals suggesting that one can be 95% confident that \( \theta \) is contained within the interval. On the one hand such problems are interpretive and could be fixed by taking researchers to task when \( p \)-values or confidence intervals are misinterpreted or the claim that more stars equate to more importance. But this will not fix the problem that frequentist inference cannot yield usable insights for many problems quantitative social science if the underlying theory is respected. Instead, the widespread and nearly irresistible tendency to misinterpret frequentist inference points to its inability to answer the questions of interest to researchers.
A.3.3 Summary

This section gave a critical overview of the logic of frequentist inference and provided two major objections to its use. Given the considerations raised above, I conclude by proposing that for many applied problems, frequentist inference cannot provide *credible* answers. In the discussion of the Berk-Freedman heuristic I argued that it is crucial to be reflective about the strengths and limitations of a particular sample. As is currently practiced, the unfettered use of *p*-values regardless of sampling considerations can only damage quantitative sociology and has a high potential to mislead. And, much more fundamentally, a statistical epistemology that centered on atheoretical mechanistic thinking giving weight to *p*-values less than 0.05 (or some other arbitrary value) while ignoring all else severely limits the sociological imagination. Fortunately there are alternatives to frequentist orthodoxy, discussed in the section of the Bayesian statistical paradigm further below. This latter perspective, in my view, provides a much more realistic, intuitive, robust, and generally useful approach for quantitative sociology.

A.4 The Assumption of Linearity

Even though frequentist inference dominates empirical sociology it has occasionally come under criticism. Even more ubiquitous—and probably equally damaging—is the assumption of linearity which is rarely, if ever, rigorously problematized. The central argument developed in this section is that treating continuous variables as linear requires an *extremely* strong and often incorrect set of assumptions. Instead it is preferable to begin with a more flexible parameterization of continuous variables, and to assume a linear relationship only if it is justified.
A.4.1 Linearity in Generalized Linear Models

Linear statistical relationships refer to a random variable $y$ that monotonically increases or decreases by the same quantity across the entire range of an explanatory variable $x$. Despite its apparent simplicity the concept of linearity is used in relatively ambiguous ways in sociological discourse. It is relatively common, for example, to see incorrect references to “nonlinear” models such as the logit or Poisson regression models. It is useful then to begin with a brief discussion of linearity within the context of generalized linear models.

Generalized linear models (GLMs) were proposed by Nelder and Wedderburn (1972) who developed an Iteratively Reweighted Least Squares (IRLS) algorithm applicable to the exponential family of probability distributions.\(^4\) GLMs consist of specifying a conditional distribution for the random variable $y$ with expectation $\mu$ given a matrix of predictor variables $X$. Since the exponential family covers several different statistical distributions, the range of $\mu$ can vary considerably. With a binary $y$, for instance, the conditional mean is constrained to the unit interval but can range between 0 and $\infty$ if $y$ is gamma-distributed. Nelder and Wedderburn’s (1972) contribution was to propose a function linking the conditional mean of $y$ to a linear predictor $\eta$, such that $g(\mu) = \eta$ and $g^{-1}(\eta) = \mu$. The linear predictor stems from the model

$$\eta = X\beta,$$

where $\theta$ is a location parameter, $a(\phi)$ is a scale parameter, and $c(y, \phi)$ is a normalizing constant.

\(^4\) The exponential family includes several of the most commonly used distributions, including the binomial, Bernoulli, Poisson, Gaussian, or gamma distributions. If the observations are identical and independently distributed then the log-likelihood of an exponential family distribution is

$$L(\theta, \phi|y_1 \ldots y_n) = \sum_{i=1}^{n} \left( \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right),$$

where $\theta$ is a location parameter, $a(\phi)$ is a scale parameter, and $c(y, \phi)$ is a normalizing constant.
which minimizes

\[ \min_{\beta} \| y - X\beta \|^2. \]  

(A.10)

Aside from Gaussian-distributed specifications, closed form solutions do not exist for GLMs. The IRLS algorithm therefore uses the linear specification of the model to iteratively estimate

\[ \beta^{(q)} = (X^TX)^{-1}X^TZ, \]  

(A.11)

where at iteration \( q \), \( Z^{(q)} = n^{(q)} + (y - \mu^{(q)}) \left( \frac{\partial n}{\partial \mu} \right) \). Because \( E(y - \mu) = 0 \) if the conditional mean is correctly specified, iterations cease when the parameter vector changes only slightly (in practice, usually \( 1E - 06 \)). At convergence the IRLS estimates are a quadratic approximation of the maximum likelihood estimates.

The discussion above clarifies the ambiguity in references to nonlinear models. The key point is that despite following nonlinear probability distributions, GLMs are based on a transformation of the linear relationship between \( X \) and \( y \) based on the parameter vector \( \beta \). References to nonlinear logistic regression models are therefore incorrect if not understood in this specific context.

### A.4.2 Accounting for Nonlinearity Using Global Corrections

When social scientists test for potential nonlinearity, the general strategy involves the introduction of polynomial transformations as regressors into a model. If the highest order polynomial term is deemed statistically significant, the relationship is considered to be nonlinear. If the polynomial is insignificant, the relationship is assumed to be linear. Polynomial regression can be useful as a rough check of a small number of nonlinear relationships, and as I discuss below, are impor-
tant components of spline models. However, there are two compelling reasons to oppose the use of simple polynomial transformations as the sole diagnostics for nonlinearities: First, simple polynomial transformations cannot account for complex data generating processes. Second, simple polynomial transformations forces the very strong assumption that a global transformation of a regressor is appropriate. If unfounded, violating this assumption can result in considerable imprecision in regression analysis. I now discuss each of these objections in more detail.

A.4.2.1 Modeling Complex Data Generation Processes

Polynomial representations in GLMs involve the transforming a predictor variable using one or more polynomial terms. If $\eta_i$ refers to the linear predictor of a response variable then the $\lambda$th degree polynomial regression model can be written as

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i1}^2 + \ldots + \beta_j x_{i1}^\lambda + \ldots + x_i^T \beta.$$  \hspace{1cm} (A.12)

Implicit in this specification is the idea that nonlinearities and polynomial representations are synonymous. It is unclear why such coarse coupling exists given the rich diversity of potential functional relationships. In practice one is generally limited to relatively lower order polynomial terms since $X$ rapidly approaches (but never arrives at) rank deficiency making coefficients and standard errors difficult to calculate precisely. This further limits the usefulness of polynomial terms to only a handful of specifications.

A more troubling consideration relates to the tendency of polynomial regression models to misdiagnose complex functional forms. Consider the following data
generating process for some $y_i$ of interest (I borrow the function from Wood (2006)):

$$y_i = ((x_i^{11}) \times ((10 \times (1 - x_i))^6)) + 10 \times ((10 \times x_i)^3) \times ((1 - x_i)^{10}) - 1.4 + \epsilon_i, \quad (A.13)$$

where $\epsilon_i$ is assumed to follow a normal distribution and $x_i$ is drawn from a uniform distribution constrained to the unit interval. Using linear regression analysis I estimated two separate models to capture this relationship ($n = 1000$). The first model contains a squared term for $x_i$ and the second has a cubic term. The Z-score for the highest order polynomial in each model is respectively $-40.03$ and $-14.376$, both of which are above conventional levels of statistical significance (to put it mildly). Based on this one could reasonably conclude that there is strong evidence of nonlinearity—which is true—that can be approximated by a second or third order polynomial—which is untrue. Neither the squared or cubic specifications come close to accurately representing $y$ making the apparent evidence for either model extremely misleading. To see this in more detail, plots of the true function and the polynomial approximations are provided in Figure A.1.
Panel A of the figure provides the true relationship between the variables which is clearly strong but quite complex. Panel B provides the fitted values from both the quadratic and cubic specifications that are graphed along with the true function. It is clear from the figure that both polynomial models provide a very poor approximation of $y$. The true relationship has a global and local maximum with a notable dip between them but such details are missed completely by the polynomial models. Both polynomial models also predict outside of the range of $y$, particularly the cubic specification. Overall, despite their high levels of statistical significance the polynomial specifications here do more harm than good since they strengthen confidence about entirely incorrect functional forms. Mistakenly reaching such conclusions could have strong theoretical implications.

A.4.2.2 The Use of Global Transformations

Another major drawback of polynomial regression is that entering a polynomial term in a regression model entails granting the assumption that the transformation is appropriate across the entire span of a regressor (Keele, 2008). This is a very strong assumption and it is unclear why it should be routinely granted. It is possible, for instance, that only a portion of the span of $x_i$ is approximated by a polynomial, after which a linear fit is appropriate. Thresholds are also likely quite common in sociological data, but our standard toolkit make them quite difficult to identify and integrate into statistical analyses. Such details, which could be quite important, are entirely lost when a global relationship is assumed. Ruppert et al. (2003) provide another example that is displayed in Figure A.2. In this case, there is a clear threshold in the relationship when $x_i$ is greater than 0 and less than 0.5, followed by relatively extensive heteroskedasticity when $x_i \geq 0.5$. Global corrections cannot account for either of the complexities in this relationship. It is
not difficult to imagine other examples relevant to sociological analyses.

The inability of global corrections to account for the selective heteroskedasticity in Figure A.2 highlights another major objection: the calculation of uncertainty around the parameters. Following the authoritative work by McCullagh and Nelder (1989), the asymptotic covariance matrix for generalized linear models is

\[ \Sigma_{\beta} \sim \phi (X^T W X)^{-1}, \]  

(A.14)

where \( W \) is a diagonal weight matrix, whose non-zero elements are contingent on the distribution of the response variable in question. In this context, \( \phi \) is treated as the dispersion parameter. For certain exponential distributions (e.g. binomial, Poisson), \( \phi = 1 \), while for others it is estimated from the data. For example, if \( y \) follows a normal distribution, then \( \phi = (n - k - 1)^{-1} \times \epsilon^T \epsilon \). Assuming that the conditional mean of the model is correct, this is equivalent to the Hessian matrix.
of the maximum likelihood estimator. The standard errors are then

$$\text{SE}(\beta) = \text{diag}(\Sigma_{\beta})^{1/2}. \quad (A.15)$$

It is clear from Equation (A.15) that assuming a global relationship exists also extends to the uncertainty around the estimates. That is, confidence intervals for polynomial regression models are insensitive to issues such as data sparsity in certain regions of a function or higher levels of variability concentrated in specific areas. Ideally, a model should be able to account for such commonalities in sociological data, but do so in a way that is not overly sensitive to local variability.

### A.4.3 Generalized Additive Models

The discussion above points to the potential problems that can come from either ignoring nonlinearities or uncritically “correcting” them using global transformations. It is desirable to use a class of models flexible enough to capture nonlinearities, that can adjust uncertainties around such estimates based on the available data, and remain interpretable. I now turn to a broad and extensible class of models that provide a much improved way to deal with complex statistical relationships using generalized additive models (GAMs). Similar to how GLMs extend the linear regression model to account for a wide variety of response variables, GAMs extend the flexibility of frequently used regression models to account for more complex relationships between predictor and response variables. Introduced authoritatively by Hastie and Tibshirani (1990), GAMs are widely used in many disciplines but re-

---

5 I proceed by assuming that one is using an IRLS estimator. For maximum likelihood estimation (MLE) the results are equivalent and therefore the same comments apply. In the case of MLE, one simply implements Newton-Raphson algorithm on the observed Hessian matrix of the log-likelihood. The coefficient covariance matrix can then be calculated directly. Hardin and Hilbe (2007, pp. 19-35), for example, provide a thorough overview of such procedures.
main relatively race in the social sciences (but see Keele, 2008; Andersen, 2009). GAMs allow for much more reasonable assumptions and much more rigorous investigations of potential nonlinearities.

To begin the exposition, which is largely drawn from Wood (2006), consider the case of model with $k$ predictor variables

$$\eta = f(x_1, x_2, \ldots, x_k). \tag{A.16}$$

For simplicity I assume that $k = 1$ and that $x_1$ is mean-centered making an intercept term unnecessary. The model in Equation (A.16) is therefore $\eta = f(x_1)$. Standard GLMs will use the IRLS algorithm in Equation (A.11) to represent the relationship between $\eta$ and $f(x)$ by a single coefficient $\beta_1$. GAMs use a more flexible representation of the relationship by introducing $k$ known basis functions, $b_j(x_1)$, allowing the model to be written as

$$f(x_1) = \sum_{j=1}^{k} b_j(x_1) \beta_j, \tag{A.17}$$

where $\beta_j$ is an unknown parameter to be estimated. The basis function transforms the predictor variables to estimate their relationship with $y$. A wide variety of basis functions are possible, ranging from the polynomials discussed above to smoothing splines. Smoothing splines provide a flexible means to represent nonlinear relationships by emphasizing local relationships, ultimately allowing changes in directionality, trajectory, and uncertainty across the range of a predictor variable. The emphasis on local rather than global transformations differs significantly from the simple polynomial transformations discussed above and therefore solves the problems that may arise from assuming a global relationship applies across the
span of $x$.

Several different splines are available, including cubic splines, natural splines, B-splines, or thin-plate splines. Splines can be represented, either directly or by approximation, as a $k$ dimensional vector $\kappa$ whose elements contain knots which serve as inflection points. For example, Wood (2006, pp. 152-153) and Keele (see 2008, pp.58-59) provide overviews of B-splines with $k$ knots, where for $i = 1 \ldots k$ the basis functions are

$$B_i^m (x) = \frac{x - \kappa_i}{\kappa_{i+2} - \kappa_i} B_i^{2-1} (x) + \frac{\kappa_{i+2} - x}{\kappa_{i+2} - \kappa_{i+1}} B_{i+1}^{2-1} (x),$$

and

$$B_i^{-1} = \begin{cases} 
1 & \text{if } \kappa_i \leq x < \kappa_{i+1} \\
0 & \text{otherwise}
\end{cases}.$$

An immediate question raised by the B-splines as well as smoothing splines more generally relates to the selection of knots. It is clear that more knots allow for greater flexibility in the specification of local relationships over the span of $x$. At the same time, one runs the risk of overfitting the spline by including superfluous knots that capitalize on patterns in the data that are noise. Consistent with the Ockham heuristic, a balance fit and parsimony is required. Fortunately, the thin plate smoothing spline simultaneously addresses both of these concerns. The underlying mechanics of thin plate splines are prohibitively complex for the current purposes (see Wood, 2006, pp. 154-160), but the central idea is to estimate a knot for each unique value of $x$. Such a procedure is clearly computationally intensive, so to minimize overfitting a smoothing parameter, $\lambda$, is introduced to penalize

---

6 In practice thin plate splines can often be represented approximately with a user-specified number of knots—usually between 10 and 20—to approximate the smooth (Ruppert et al., 2003), though this comes at the loss of certain optimality properties (Wood, 2006).
highly nonlinear functions. This is achieved by constructing the roughness penalty

\[ \lambda \int_a^b (f''(x))^2 \, dx, \]

which penalizes highly nonlinear functions that have large squared second derivatives. This roughness penalty is added to the estimating equation, making the IRLS algorithm in Equation (A.10) and Equation (A.11) applicable to models containing smooth terms. GAMs can therefore be considered to be penalized implementations of the IRLS algorithm where the goal is to minimize

\[ \min_{\beta, f(x)} \| Y - X\beta - f(x) \|^2 + \sum_{j=1}^J \lambda_j \int_a^b (f''(x_j))^2 \, dx_j. \]  

(A.18)

Note that it is implicit in Equation (A.18) that when a linear fit between \( x \) and \( y \) is appropriate, this will be reflected in the smooth.

To put the various pieces of this discussion together, the GLM notation above coupled with this discussion of smoothing can be used to represent GAMs as arbitrary mixtures of parametric coefficients and smooths. In equation format, a model containing both parametric estimates and multiple smooths can be written as

\[ \eta = X\beta + f(x). \]

Such a specification is quite broad allowing for the inclusion of categorical variables or interaction terms (even between smooths) and is applicable to any member of the exponential family. Further extensions, such as generalized additive mixed models, are also possible (Wood, 2006).
A.4.4 Summary

This section provided a critique of either ignoring nonlinearities or uncritically representing such relationships using simple polynomial transformations. As a corrective, I proposed using generalized additive models which rely on local relationships to provide a flexible representation of nonlinearity through the use of smoothing splines. Despite their advantages GAMs are far from a panacea. They come with a heavy computational cost and are generally more difficult to interpret than a single regression coefficient. Such concerns are offset by providing a more honest representation of the relationship between two variables. GAMs are also criticized for consuming far more degrees of freedom compared to their GLM counterparts. Given the discussion of the Ockham heuristic above, such concerns are ill-warranted. If retaining degrees of freedom is not taken as an end in and of itself then the simplicity should not be privileged over fit, particularly when it can lead to considerable misspecification of functional forms.

A.5 The Bayesian Paradigm

In 1764, the Reverand Thomas Bayes’ “Essay Towards Solving a Problem in the Doctrine of Chances” was posthumously published in *Philosophical Transactions of the Royal Society of London*. The essay provided an innocuous law used in the calculations of conditional probabilities. Bayes’ proposal demonstrated that

$$\Pr (B|A) = \frac{\Pr (A|B) \Pr (B)}{\Pr (A)}, \quad (A.19)$$

where $\Pr (B|A)$ is the posterior probability, $\Pr (A|B)$ is the likelihood, $\Pr (B)$ is the prior distribution, and $\Pr (A)$ is the probability of the data. The posterior
probability is used for inference in the Bayesian paradigm. In Equation (A.19) $B$ is a unknown parameter while $A$ refers to fixed data. While there are some comparabilities to frequentist reasoning here (e.g. the likelihood remains unchanged), Bayes’ proposal reverses the logic of what is fixed and what is random: from a Bayesian perspective the data is fixed and the parameter is random, while the opposite is true for frequentists. This reversal has the considerable advantage to provide direct probabilistic estimates of the value of $B$ given the data without reference to an imagined sampling distribution of statistics. Bayes theorem, at least in principle, therefore provides a useful inferential engine that is more intuitive to how researchers reason (at least given the consistent manner in which frequentists use Bayesian interpretation).

By far the most contentious component of Bayes theory involves the necessary specification of a prior distribution. A thorough discussion and review of the literature on priors is beyond my scope, though a few short comments are in order. In practice, a prior refers to the analyst’s set of beliefs about a parameter vector of interest that is external to the data. Such beliefs are translated into probability distributions and entered into Bayes Theorem. Priors are wholly subjective constructs and strong priors can considerably change the posterior probabilities. For this reason, some have objected to the Bayesian approach entirely as too subjective for “serious” research, which (falsely) purports to be a wholly objective enterprise. It is true that priors can be abused, but such considerations are generally misguided if not entirely incorrect. All statistics models have elements of subjectivity and pretending that no prior knowledge exists is no more objective than including it systematically. Bayesian models are also not unique in their ability to be manipulated by less than scrupulous analysts. The dirty secret of frequentist modeling is that many analyses contain only a snapshot of the dozens or hundreds of
models that were run until a pattern of significant coefficients is deemed suitably interesting for publication. Gill (2008, 135-185) provides a useful categorization that distinguishes between conjugate, informative, and uninformative priors. The latter are by far the most often used in the rare applications of Bayesian methodology in the social sciences, though conjugate priors (e.g. beta-binomial models) are occasionally used as well. Informative priors are (unfortunately) rarely used.

Putting the discussion of priors aside, a more immediate complication apparent from Bayes Theorem arises when one attempts to analytically evaluate the denominator. Solving for the posterior probabilities using Bayes theorem involves summing or integrating across the entire parameter space of $B$. When the denominator of Bayes theorem refers to a continuous density, Equation (A.19) is rewritten as

$$
\Pr (B|A) = \frac{\Pr (A|B) \Pr (B)}{\int \Pr (A|B) \Pr (B) \, dB},
$$

and when the denominator refers to a categorical density it is

$$
\Pr (B|A) = \frac{\Pr (A|B) \Pr (B)}{\sum \Pr (A|B) \Pr (B)}.
$$

When $B$ refers to multiple parameters, analytically solving Bayes theorem rapidly becomes anywhere from difficult to impossible, even with current computing power. As a consequence, the vast majority of applied Bayesian problems are intractable. Despite the potential complexity of the denominator in Bayes theorem, it is clear from Equation (A.19) that it is constant across $1 \ldots n$. The denominator can therefore be omitted from the equation, allowing us to further rewrite Bayes theorem by solving up to a constant of proportionality. In equation format, this is

$$
\Pr (B|A) \propto \Pr (A|B) \Pr (B).
$$
This latter equation forms the basis of what Jackman (2009) refers to as the Bayesian mantra: the posterior is proportionate to the likelihood times the prior. As is clear from Equation (A.19), the constant of proportionality is simply \(1 / \Pr(A)\).

The workaround to avoid analytically evaluating the denominator in Bayes theorem still leaves a major complication. In multiparameter problems the joint posterior distribution can still be extraordinarily difficult to evaluate. Joint densities above two dimensions cannot be easily examined graphically nor are there useful numerical summaries. Put another way, we are generally interested in the marginal posterior distributions for each parameter of interest, and a \(k\)-dimensional posterior density requires integrating over \(k - 1\) distributions. Consider the following example: Assume that \(Y \sim \mathcal{N}(X \hat{\beta}, \sigma^2 I)\), that \(X\) is mean-centered, and that

\[
\eta = \hat{\beta}_1X_1 + \hat{\beta}_2X_2 + \hat{\beta}_3X_3 + \hat{\beta}_4X_4 + \hat{\beta}_5X_5 + \hat{\beta}_6X_6.
\]

If we are interested in the marginal posterior probability of, say, \(\hat{\beta}_2\), then we must solve

\[
\Pr\left(\hat{\beta}_2 | \hat{\beta}_1, \hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5, \hat{\beta}_6, \sigma^2, A\right) \propto \int_{\hat{\beta}_1} \int_{\hat{\beta}_3} \int_{\hat{\beta}_4} \int_{\hat{\beta}_5} \int_{\hat{\beta}_6} \int_{\sigma^2} \Pr\left(A | \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5, \hat{\beta}_6, \sigma^2\right) \\
\times \Pr\left(\hat{\beta}_1\right) \times \Pr\left(\hat{\beta}_2\right) \times \Pr\left(\hat{\beta}_3\right) \times \Pr\left(\hat{\beta}_4\right) \\
\times \Pr\left(\hat{\beta}_5\right) \times \Pr\left(\hat{\beta}_6\right) \times \Pr\left(\sigma^2\right).
\]

(A.20)

Clearly this is far from a trivial task. Fortunately a robust literature exists allowing for the evaluation of complex integrals using Monte Carlo integration, and now problems of high complexity can be straightforwardly approximated using numeri-
cal computing. Such algorithms form the basis of applied Bayesian estimation and I now turn to a more detailed discussion of their functioning.

A.5.1 Estimation by Markov Chain Monte Carlo

Markov chain Monte Carlo (MCMC) methods are perhaps one of the central reasons for the resurgence of interest in Bayesian methodology. The central idea is to use Markov chains to generate a stochastic process allowing for Monte Carlo integration of complex parameter spaces. Contemporary Bayesian analyses are based on simulations which provide entire posterior densities. Gilks et al.’s (1996) widely cited edited volume *Markov Chain Monte Carlo in Practice* provided an accessible way to estimate extremely complex Bayesian models, and MCMC methodology remains an active area of interest and research (e.g. Brooks et al., 2010). To provide a skeletal overview of MCMC techniques, I first split them into their component parts by reviewing Markov chains and then Monte Carlo integration. I then bring the two techniques together by outlining the Gibbs sampler.

A.5.1.1 Markov Chains

Following Gill (2008, pp. 344-355), Markov chains are ordered stochastic processes occurring over the range of a parameter space $\Theta$, or more simply a series of random quantities occurring within the allowable range of $\Theta$. The key property of Markov chains for the current purposes is that given a value $\theta^{t+1}$, the probability of transition to to a new state in the parameter space given $\theta^t$ does not depend on previous values $\theta^1, \ldots, \theta^{t-1}$. In equation format, given that $A \subset \Theta$ represents a subset of the entire parameter space, this can be summarized more generally as

$$\Pr \left( \theta^{t+1} \in A | \theta^1, \theta^2, \ldots, \theta^{t-1}, \theta^t \right) = \Pr \left( \theta^{t+1} \in A | \theta^t \right).$$
In other words, again borrowing from Gill (2008), a Markov chain will traverse the parameter space and is hindered only by the previous values. Proofs exist establishing that Markov chains will converge towards their stationary distribution, a condition where transferring from one state to another retains the same marginal distribution. I discuss the issue of convergence in more detail by working with an example below. It is important to note, however, that upon convergence, the values of the Markov chains provide the posterior densities that are the key to Bayesian inference.

The principles underlying Markov chains can be easily illustrated by example, in this case using a discrete Markov process (other useful examples can be found in Albert, 2007; Gill, 2008). Consider a case where five cards, labeled 1 through 5, are aligned sequentially. Once an initial card is selected, one can stay in the same location or move one value to the left or right for values 2, 3, and 4. For border values (i.e. 1 and 5) once can either stay in the same location, or move to the directly adjacent number (2 and 4 respectively).\(^7\) For simplicity, it is assumed that there are no preferences of one number over another. Putting these together yields the following transition matrix:

\[
T = \begin{bmatrix}
\frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\
0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\
0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
0 & 0 & 0 & \frac{1}{2} & \frac{1}{2}
\end{bmatrix}.
\]

Prior to starting the algorithm, it is also necessary to choose starting values at

\(^7\) Note that this example is uses a random walk process, which can be used for priors when dealing with nonparametric estimates in a Bayesian model.
iteration 1. In this case, I assume that each of the five cards have equal probability of selection. The elements of the vector of starting values are each 1/5, giving

\[ p_1^T = \begin{bmatrix} 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \end{bmatrix}. \]

An important property of Markov chains (and MCMC algorithms more generally) is that similar to maximum likelihood estimators, convergence is not contingent on the choice of starting values. The number of iterations required for convergence may change, however, as the number of iterations approach infinity, the results are indistinguishable. To demonstrate this, I define a second vector of starting values:

\[ q_1^T = \begin{bmatrix} 1/10 & 1/5 & 1/2 & 1/10 & 1/10 \end{bmatrix}. \]

| Table A.1. Markov chain iterations for two different starting values |
|-------------------|---|---|---|---|---|
| Iteration | \( p_1 \) | \( p_2 \) | \( p_3 \) | \( p_4 \) | \( p_5 \) |
| Starting Values 1 |
| 1 | 0.20 | 0.20 | 0.20 | 0.20 | 0.20 |
| 2 | 0.17 | 0.23 | 0.20 | 0.23 | 0.17 |
| ... | | | | | |
| 1000 | 0.15 | 0.23 | 0.23 | 0.23 | 0.15 |
| Starting Values 2 |
| 1 | 0.10 | 0.20 | 0.50 | 0.10 | 0.10 |
| 2 | 0.15 | 0.24 | 0.27 | 0.21 | 0.13 |
| ... | | | | | |
| 1000 | 0.15 | 0.23 | 0.23 | 0.23 | 0.15 |

Having defined the transition matrix and two different vector of starting values, it is simply a matter of calculating \( p_i^T T \) over \( i = 2 \ldots N \) iterations until the algorithm has converged to its stationary distribution. This occurs when the marginal distribution remains unchanged such that \( p^T T = p^T \). Continuing with the example, a select number of iterations are provided for both sets of starting values in
in Table A.1. As can be seen in the table, both sets of starting values converge so that \( p^T T = q^T T \) and \( p = q \). Given a sufficient number of iterations, Markov chains will converge to their stationary distribution regardless of the complexity of the transition matrix and parameter space.

### A.5.1.2 Monte Carlo Integration

Monte Carlo integration is a numerical technique which uses pseudo-random numbers to evaluate integration problems (Robert and Casella, 2004). As Jackman (2009, p. 133) notes, the Monte Carlo principle is that “anything we want to know about a random variable \( \theta \) can be learned by sampling many times from \( f(\theta) \), the density of \( \theta \).” Put formally, Gill (2008, pp. 277–279) demonstrates that if we are interested in the distribution of the probability function \( g(\theta) \) over the domain \([a,b]\) we must solve

\[
E(h(\theta) \in [a,b]) = \int_a^b g(\theta) h(\theta) \, d\theta.
\]

We can then invoke the Monte Carlo principle and generate \( T \) samples so that

\[
E(h(\theta) \in [a,b]) = \frac{1}{T} \sum_{t=1}^{T} h(\theta).
\]  

(A.21)

Using the strong law of large numbers, it follows that as \( T \to \infty \), \( E(h(\theta) \in [a,b]) \) converges to \( E(h(\theta) \in [a,b]) \).\(^8\) Similar comments apply to the variance and higher moments about the mean as well. Recast in terms of Bayes theorem, if we let \( g(\theta) \) refer to the posterior probability \( \text{Pr}(\theta | A) \) then Gill (2008) shows that Equa-

\(^8\) As I mentioned above, several frequentist properties are widely used in Bayesian methodology. Here we see another example where an theoretically limitess number of draws is used to draw conclusions about a probability distribution of interest, which is similar in logic to frequentist inference. The difference is that Monte Carlo integration provides the analyst with \( T \) random draws so one can still avoid making reference to purely theoretical sampling distributions.
tion (A.21) can be written as

\[ E(\theta|A) = \int \Pr(\theta|A) h(\theta) d\theta, \]

and therefore the posterior mean is

\[ E(\theta|A) \approx \frac{1}{T} \sum_{t=1}^{T} h(\theta). \]

Since Bayesian models generally provide complex posterior densities that often have an unknown form—the exception being when conjugate priors are used—Monte Carlo techniques become incredibly useful. Regardless of the complexity of the underlying distribution, when enough samples are drawn Monte Carlo integration can provide all the information that is needed for analysis and inference. Having outlined the logic of both Markov chains and Monte Carlo methods, I now turn to a discussion of the Gibbs sampler, a widely used algorithm that provides a synthesis of both techniques.

A.5.1.3 The Gibbs Sampler

The Gibbs sampler is a workhorse Bayesian algorithm that is used in a wide variety of applications.\(^9\) Gibbs sampling simplifies the calculation of conditional and marginal posterior densities using the advantageous properties of Markov chains and Monte Carlo methods described above. As a result, Gibbs sampling makes the joint posterior density available even if it is analytically intractable. An introduction to Gibbs sampling is provided by Casella and George (1992), and more detailed

\(^9\) This is reflected, for example, in the widely used Bayesian software packages WinBUGS, OpenBUGS, and JAGS. BUGS is an acronym for Bayesian Inference Using Gibbs Sampling while JAGS stands for Just Another Gibbs Sampler.
discussions are provided by Lynch (2007), Gill (2008), and Jackman (2009).

To provide an overview of how the Gibbs sampler operates, let $\theta$ refer to a $k$-dimensional vector of parameters that form an unknown $k$-dimensional joint density $f(\theta_1, \theta_2, \ldots, \theta_k)$. Accessing the marginal distribution for, say, $\theta_1$ requires solving a rather formidable multiple integral. In contrast, the Gibbs sampler is based on a Markov chain over $T$ draws where at each iteration the algorithm alternates between $\theta_1$ through $\theta_k$, conditional at the other values at iteration $t-1$. The value for $\theta^{t-1}$ is substituted for $\theta^t$ as the algorithm cycles through each parameter of interest. This can be seen more clearly in equation form, where at iteration $t$ the Gibbs sampler is

$$
\begin{align*}
\theta_1^t & \sim \Pr \left( \theta_1 \mid \theta_2^{t-1}, \ldots, \theta_k^{t-1} \right) \\
\theta_2^t & \sim \Pr \left( \theta_2 \mid \theta_1, \ldots, \theta_k^{t-1} \right) \\
& \vdots \\
\theta_k^t & \sim \Pr \left( \theta_k \mid \theta_1, \ldots, \theta_{k-1} \right)
\end{align*}
$$

Once the Markov chain has converged to its stationary distribution, we can rely on the Monte Carlo principle to sample from the marginal density of our choosing. Gibbs sampling therefore uses a ‘divide-and-conquer’ strategy to solve intractable integration problems. Since the Gibbs sampler converges to the true underlying distribution if $T$ is sufficiently large, it can be used to tell us any desired information about the distribution of each parameter. In the case of a Bayesian analysis then, it is clear how the Gibbs sampler allows us to make direct probabilistic inference about a parameter given the data since we have effectively created its sampling distribution through simulation.
A.5.2 Summary

This section provided a brief overview of Bayesian reasoning and estimation. This literature has vastly expanded over the last two decades making anything but the most tertiary treatment a monumental task. I have omitted discussion of widely used estimation procedures (e.g. Metropolis-Hastings algorithms) as well as promising new developments such as MCMC estimation using Hamiltonian dynamics proposed by Neal (2011). I also avoided other key components of Bayesian methodology including Markov chain convergence, sensitivity analyses, and Bayes factors. These omissions, I hope, do not overshadow the considerable advantages and flexibility that Bayesian analysis can bring to the table. In essence, if it is possible write a likelihood function for a model and one can assign priors, Bayesian approaches can provide any quantity of interest with excellent optimality properties and provides a much more realistic and intuitive way to make inference. The use of priors also provides a ripe avenue to more systematically create a symbiotic relationship between qualitative and quantitative research.

A.6 General Analytic Strategy

This final section briefly synthesizes the discussion above by outlining the Bayesian semiparametric mixed logistic regression models used in Chapters 2 through 4. I believe that these models provide a flexible that is honest to the underlying data patterns, while respecting the core components of the Box Heuristic, the Berk-Freedman Heuristic, and the Ockham Heuristic. Before outlining the algebra of these models, I sketch out one more useful equivalency to establish that certain specifications of GAMs can be treated as special cases of mixed models. In the case of Bayesian analysis, fitting GAMs as mixed models dramatically simplifies
the estimation procedure.

### A.6.1 Formulating GAMs as Mixed Models

As I demonstrated above in Equation (A.18), GAMs are penalized versions of the iteratively reweighted least squares algorithm. If one is willing to represent smooths using a finite number of knots then Ruppert et al. (2003, pp. 65-66; 108-110) prove that the estimating equations can be rewritten as mixed models, and in fact provide the best linear unbiased predictor if \( y \) is Gaussian. The proofs are quite involved and I avoid them here, however the end result is that it is possible to rewrite the smooth of \( x \) as

\[
f(x_i) = \beta_0 + \beta_1 x_i + \sum_{k=1}^{\kappa} b(x_i - \kappa_k) \beta.
\]

Then we can define the usual matrix of observed coefficients \( X \) as

\[
X = \begin{bmatrix}
1 & x_1 \\
\vdots & \vdots \\
1 & x_n
\end{bmatrix}
\]

and a \( n \times \kappa \) matrix of deviations from the knot points \( Z \) as

\[
Z = \begin{bmatrix}
x_1 - \kappa_1 & \ldots & x_1 - \kappa_k \\
\vdots & \ddots & \vdots \\
x_n - \kappa_1 & \ldots & x_n - \kappa_k
\end{bmatrix}.
\]

Putting the pieces together allows us to write the mixed model

\[
y = X\beta + Zb + \epsilon_{ij}. \quad (A.22)
\]
It is assumed in Equation (A.22) that \( y \) is drawn from a normal distribution (though this is unnecessary). The smoothing spline can therefore be represented using normally distributed random effects denoted by \( b \), while \( \beta \) remains the vector of fixed coefficients. This representation can be extended further to include an arbitrary number of smooths and fixed coefficients, limited and categorical outcomes as well as multiple levels within the mixed model (Crainiceanu et al., 2005).

### A.6.2 Bayesian Semiparametric Mixed Models

Mixed models are particularly attractive from a Bayesian perspective because more aggregate units provide natural prior information that can be incorporated into the model (Gelman and Hill, 2007), as we will see momentarily. They also have excellent optimality properties that generally provide better performance relative to Penalized Quasi-Likelihood or sixth-order LaPlacian estimation procedures which provide only approximation of the likelihood (Diaz, 2007; McCulloch and Searle, 2001; Browne and Draper, 2006).

To outline the Bayesian representation of the mixed model, let \( \beta \) be a \( K \)-dimensional vector containing all fixed coefficients and define \( b \) as a \( Q \)-dimensional vector of random effects. As above, \( X \) is a matrix of observed variables and \( Z \) is a composite of identification variables used to map random effects and transformed values of \( x \) for smooths. \( \Sigma \) is a positive-definite covariance matrix of the random effects and potentially the level-one error variance if \( y \) is Gaussian. The joint posterior density for Bayesian mixed models, up to a constant of proportionality, is

\[
\Pr (\beta, b, \Sigma \mid X, Z, y) \propto L (X, Z, y \mid \beta, b, \Sigma) \Pr (\beta) \Pr (b \mid \Sigma) \Pr (\Sigma)
\]

Note that the above equation conditions the prior of \( b \) on its covariance matrix \( \Sigma \).
This is referred to as a hyperprior in the Bayesian literature, which links aggregate and lower level units as I mentioned above. Once the likelihood function for a two-level logistic regression model is substituted, and assuming a Bernoulli distributed for $y$, the final model specification written in expanded form is the rather formidable

$$
\text{Pr} (\beta, b, \Sigma | X, Z, y) \propto \prod_{j=1}^{J} \left( \prod_{i=1}^{n} \left( \left[ 1 + \exp (x'_{ij}\beta + z'_{ij}b) \right]^{-1} \right)^{y_{ij}} \right)
\times \left( 1 - \left[ 1 + \exp (x'_{ij}\beta + z'_{ij}b) \right]^{-1} \right)^{1-y_{ij}} db
$$

(A.23)

Equation (A.23) provides the baseline model used in Chapters 2, 3, and 4.

A.6.3 Conclusion

This appendix outlines the logic underpinning my analytic strategy. This modeling strategy is fairly complex and clearly only one of many legitimate approaches, but I believe it provides a flexible approach to the substantive issues addressed in this project. Another advantage is that it is remarkably straightforward to program (see Crainiceanu et al., 2005). To close, it is worthwhile to briefly dispel the inevitable criticism that this approach rests on overfitting or overcomplicating the statistical analyses. I believe that my best defense is the substantive pattern of results. Chapter 2, for example, demonstrates a remarkable departure from linearity for the effect of protest size on the likelihood of arrest. Similar comments apply to the smooth of media attention on surveillance patterns in chapter 4. The use of newspaper data also makes frequentist inference either very difficult or impossible to evaluate. Overall then, I believe my methodology is at least defensible.
# Cities and Newspaper Sources

## Table B.1. List of Cities and Newspaper Sources

<table>
<thead>
<tr>
<th>City</th>
<th>Newspaper Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atlanta</td>
<td>Atlanta Journal-Constitution</td>
</tr>
<tr>
<td>Austin</td>
<td>Austin American Statesmen</td>
</tr>
<tr>
<td>Boston</td>
<td>Boston Globe</td>
</tr>
<tr>
<td>Chicago</td>
<td>Chicago Tribune</td>
</tr>
<tr>
<td>Columbus</td>
<td>Columbus Dispatch</td>
</tr>
<tr>
<td>Denver</td>
<td>Denver Post</td>
</tr>
<tr>
<td>Detroit</td>
<td>Detroit Free Press</td>
</tr>
<tr>
<td>Indianapolis</td>
<td>Indianapolis Star</td>
</tr>
<tr>
<td>Jacksonville</td>
<td>Florida Times-Union</td>
</tr>
<tr>
<td>Louisville</td>
<td>The Courier-Journal</td>
</tr>
<tr>
<td>Miami</td>
<td>Miami Herald</td>
</tr>
<tr>
<td>Milwaukee</td>
<td>Milwaukee Journal-Sentinel</td>
</tr>
<tr>
<td>Nashville</td>
<td>The Tennessean</td>
</tr>
<tr>
<td>New York</td>
<td>New York Times</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>Philadelphia Inquirer</td>
</tr>
<tr>
<td>Phoenix</td>
<td>Arizona Republic</td>
</tr>
<tr>
<td>Portland</td>
<td>The Oregonian</td>
</tr>
<tr>
<td>San Francisco</td>
<td>San Francisco Chronicle</td>
</tr>
<tr>
<td>Seattle</td>
<td>Seattle Post-Intelligencer</td>
</tr>
<tr>
<td>Washington</td>
<td>Washington Post</td>
</tr>
</tbody>
</table>
Bibliography


Cunningham, David and John Noakes. 2008. ““What If She’s from the FBI?” The Effects of Covert Forms of Social Control on Social Movements.”* Sociological Science and Deviance* 10:175–197.


194


Vita
Patrick Rafail

The Pennsylvania State University
Department of Sociology
211 Osvald Tower
University Park, PA 16803

Phone: (814) 321-7877
Fax: (814) 863-7216
Email: psr139@psu.edu

Education

• M.A. Sociology and Social Statistics, McGill University, 2005.
• B.A. Sociology, McGill University, 2004.

Research and Teaching Interests
Social Movements/Collective Behavior, Applied Statistics, Organizations, Policing

Select Publications


