Title
Three Essays on Measuring Social Context in the Social Sciences

Author
Zachary, Paul

Publication Date
2018

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Three Essays on Measuring Social Context in the Social Sciences

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Political Science

by

Paul Zachary

Committee in charge:

Professor James Fowler, Chair
Professor Joshua Graff Zivin
Professor Seth J. Hill
Professor Thad Kousser
Professor Margaret E. Roberts

2018

The dissertation of Paul Zachary is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California San Diego

2018

DEDICATION

To Alan. Thank you. For everything.

EPIGRAPH

Is it possible, in the final analysis,

for one human being to achieve perfect understanding of another?

We can invest enormous time and energy in serious efforts to know another person,

but in the end, how close can we come to that person's essence?

We convince ourselves that we know the other person well,

but do we really know anything important about anyone?

—Haruki Murakami, The Wind-Up Bird Chronicle

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I could not have contemplated writing a dissertation without the endless love, help, and support from so many individuals who are too numerous to name individually here. Many times throughout this process, I found myself ready to throw in the towel and give up, only to be pulled back from the brink by the encouragement of my friends and mentors. While words cannot fully convey the scope of my gratitude, I will do my best.

I am forever in Alexander Downes' debt for mentoring me. He agreed to meet for hours each week to discuss my independent study. It is thanks to him that I have a firm sense of the literature on political violence. He also held my hand through my first publication process. He helped improve my writing and the clarity of my thoughts. It is because of him that I can call myself a social scientist today.

James Fowler is perhaps the most important mentor I've ever had. When we met for the first time during recruitment week, I immediately felt that I had found my academic "home." I did not doubt his belief in me as a scholar and as a person. This confidence allowed me to take on challenges I never would have otherwise. I will also be forever grateful to James and his wife Harla for opening their home to me when I needed a place to stay in San Diego, CA.

Finding someone whose judgment you trust instantaneously is not common in academia. I found such a person in Christopher Fariss. He meets every idea and every project — no matter how farfetched — with an infectious sense of enthusiasm. I am grateful for his encouragement and support.

Beyond academic mentors, I owe a great debt to my family and friends. My mother never got cross with me when I needed to spend all day in the library when visiting her. She helped me clean data when I needed extra hands. Her love helped me make it through this long journey. I am so grateful to her.

I also truly grateful to the mentorship of my dissertation committee members: Thad Kousser, Seth Hill, Molly Roberts, and Josh Graff Zivin. Thank you!

Last but not least is my husband Alan McLachlan. I cannot put into words how much his love and support over the years has meant to me. There were too many nights to count when self-doubt nearly overwhelmed me. He supported my passion, wherever it took me. Most importantly, he made sure I knew I had his support. The only reason I have any writing to submit is because of his support and love, every step of the way. I love you.

Chapter 2 in part is currently being prepared for submission for publication of the material. Chen, Ted Hsun Yun; Christopher J. Fariss; and Paul Zachary.

Chapter 3, in part is currently being prepared for submission for publication of the material. McCord, Gordon, Joseph Brown, and Paul Zachary.

Chapter 4, in full, is a reprint of the material as it appears in Chen, Ted Hsuan Yun, Paul Zachary, and Christopher J. Fariss. Computational Social Science 2017: Proc. 2017 Annual Conf, David Krakauer and Scott Page, Eds.

VITA

2010	B. A. in Political Science *magna cum laude*, Columbia University
2013	M. A. in Political Science, The George Washington University
2013-2018	Graduate Fellow, University of California San Diego
2018	Ph. D. in Political Science, University of California San Diego

PUBLICATIONS

Zachary, Paul, Kathleen Deloughery, and Alexander Downes. 2017. "No Business like FIRC Business: Foreign-Imposed Regime Change and Bilateral Trade," *British Journal of Political Science* 47 (4): 749-782.

Chen, Ted, Chris Fariss, and Paul Zachary. "Social Activity Hubs (SAHs): A New Method for Estimating User Specific Contextual Factors From Social Media Data," in *Computational Social Science 2017*: Proc. 2017 Annual Conf, David Krakauer and Scott Page, Eds.

Zachary, Paul and William Spaniel."Getting a Hand By Cutting Them Off: How Uncertainty over Political Corruption Affects Violence." Forthcoming, *British Journal of Political Science*.

ABSTRACT OF THE DISSERTATION

Three Essays on Measuring Social Context in the Social Sciences

by

Paul Zachary

Doctor of Philosophy in Political Science

University of California San Diego, 2018

Professor James Fowler, Chair

Most social science theories attempt to explain the ways in which social interactions affect human behavior. Inferences therefore depend upon a researchers' ability to measure and characterize these interactions. In this dissertation, we highlight several inferential challenges in measuring social interactions and show how geographical data can be used to address them. One such challenge is posed by exposure to various treatments; many treatment variables of interest are difficult to measure in a coherent and unified way. Chapter 2 illustrates the way in which public records can be used to overcome this challenge. Another challenge is post-treatment sorting. Once social conflict begins, one of the first things that groups do is segregate themselves into smaller sub -groups. Chapter 3 introduces a new instrumental variable to correct for this bias

in census data. Lastly, Chapter 4 introduces a novel method for estimating where social media users spend time during their day. These findings and novel methods will contribute to social scientists ability to study contextual effects on behavior.

1 Introduction: Measuring Context in Behavioral Social Science

1.1 Introduction

Many social science theories attempt to model how social interactions or knowledge influence human behavior. These theories make predictions about how people will behave in one social context relative to another (Kasara 2016; Habyarimana et al. 2007; Sands 2017; Enos and Gidron 2016; Enos 2014; Bhavnani et al. 2014), what they will do when interacting with one type of institution versus another (Jha 2013; Lipsky 2010; Hartzell and Hoddie 2003; Posner 2005), or why they hold certain beliefs as opposed to others (Bhavnani, Findley, and Kuklinski 2009; Berinsky 2017; Greenhill and Oppenheim 2017). As a result, the quality of empirical social science research is conditional upon scholars' ability to precisely measure the social context and interactions of individuals under study.

Obtaining such data is enormously challenging. In addition to the threats to validity posed by endogeneity and empirical strategy, poor data can negatively impact conclusion validity. For instance, consider the potential relationship between ethnic or racial diversity and conflict onset and intensity. Using national-level measures of diversity, one group of scholars finds no relationship between diversity and conflict onset (Fearon and Laitin 2003; Collier and Hoeffler 2004). Using subnational measures of diversity, another group finds a negative relationship between diversity and conflict onset and intensity (Kasara 2016). In Chapter 4, we address these conflicting findings by measuring diversity on a much more fine-grained scale (down to the block of a city). We find that diversity increases conflict intensity. As evidenced by this example, the overall conclusion is dramatically impacted by the scale at which diversity is measured (at the national, subnational, or city block level).

The ability to accurately measure behavior is crucial to the enterprise of social science research. Many of the behavioral phenomena that social scientists study (e.g., bribe-taking, protesting) are difficult to observe and measure (Zachary and Spaniel 2018). Data obtained from self-reported surveys are subject to numerous well-known biases including social desirability and framing effects (seems like you should have references here to support this statement). Newly developed data collection methods, such as analysis of data obtained from forms of social media, offer new opportunities to measure behavior. However, it is important to note that these data often lack measures of context. These measures can provide further insight by revealing contextual factors that influence behavior in multiple ways. In Chapter 3, we introduce a new method to estimate social context from social media data.

The size of the unit of analysis and associated assumptions of ecological inference are another potential source of bias in empirical studies (King 2013; King, Tanner, and Rosen 2004). When researchers make predictions about the effect of social interactions on behavior they often test these predictions by using highly aggregated units of analysis, such as census districts or wards. For example, studies on the effect of demographic change on support for far-right political parties uses census data to measure ethnicity (Dancygier 2010; Arzheimer 2009). However, using the average number of different ethnic groups that reside within a census enumeration district can obscure divergent settlement patterns. For example, a district where two ethnic groups live geographically intermixed will have the same average value as one where they are entirely segregated.

To overcome the challenge of measurement, many of the aforementioned studies rely upon some form of randomized intervention, be it in the lab, the field, or naturally occurring. In these studies, each subject's context is manipulated in various ways. For example, in one study participants encountered a homeless person (staged by the experimenters) as they completed a survey. This allowed the experimenters to measure the way in which visible poverty affects altruistic sentiment (Sands 2017). Another study engaged participants who self-identified as white/Caucasian in a game in which they were asked to play the role of a country's dictator. When these participants had been previously informed about the growth of the Hispanic population in the United States, they were more likely to favor other white participants (Abascal 2015). Research has also found that transphobia (prejudice against transgender people) is reduced when canvassers randomly encourage active-perspective-taking among survey respondents (Broockman and Kalla 2016).

Randomized experiments convincingly establish causality and can rule out alternate explanations. However, the external validity of such findings is not always clear (Fariss and Jones 2017; Shadish 2010; Shadish, Cook, and Campbell 2001). This issue is particularly highlighted in experiments that attempt to measure political and social context. By necessity, these studies tend to use extremely strong treatments in order to generate statistically detectable effects given limited sample sizes. While observational studies do not typically have strong internal validity, they have excellent external validity. Observational studies measure subject specific-behaviors in real life situations, often without direct intervention on behalf of the experimenters.

In this chapter, we briefly discuss some of the challenges encountered in prior empirical research that explores the way in which context affects political behavior. We then summarize this dissertation's contribution to the literature.

1.2 Road Map

This dissertation presents novel research designs and methods to overcome the multiple challenges described above. In Chapter 2, we study how contact with "street-level bureaucrats" affects the propensity to participate in collective action (Lipsky 2010). Specifically, we use micro-level data of arrests obtained via public records of the Baltimore Police Department to study how exposure to policing affected participation in the Freddie Gray protests in April 2015. We show that people who spent time in areas with higher police contact were more likely to engage in collective action during the riots.

In Chapter 3, we show how endogenous population sorting at the start of civil war can bias inferences about the relationship between ethnic diversity and conflict intensity. Using an instrumental variable, we correct measurement error in the 1973 United Kingdom census to model where Catholics and Protestants lived in Belfast prior to the start of the Troubles. We show that failing to correct for these population movements underestimates the effect of diversity on conflict intensity by nearly 50%.

Lastly in Chapter 4, we introduce a novel method to estimate the locations where social media users spend their time. This new tool enables researchers to use social media data to measure behavior and control for user-specific contextual factors such as socioeconomics and demographics.

To summarize this dissertation's contribution to the literature, we demonstrate the importance of properly measuring context in social science research across a variety of applications. We also introduce a novel quantitative method that permits analysis of the impact of context when using data from social media as a sampling frame.

Bibliography

Abascal, Maria. 2015. “Us and them: Black-White relations in the wake of Hispanic population growth.” *American Sociological Review* 80 (4): 789–813.

Arzheimer, Kai. 2009. “Contextual factors and the extreme right vote in Western Europe, 1980–2002.” *American Journal of Political Science* 53 (2): 259–275.

Berinsky, Adam J. 2017. “Rumors and health care reform: experiments in political misinformation.” *British Journal of Political Science* 47 (2): 241–262.

Bhavnani, Ravi, Karsten Donnay, Dan Miodownik, Maayan Mor, and Dirk Helbing. 2014. “Group segregation and urban violence.” *American Journal of Political Science* 58 (1): 226–245.

Bhavnani, Ravi, Michael G Findley, and James H Kuklinski. 2009. “Rumor dynamics in ethnic violence.” *The Journal of Politics* 71 (3): 876–892.

Broockman, David, and Joshua Kalla. 2016. “Durably reducing transphobia: A field experiment on door-to-door canvassing.” *Science* 352 (6282): 220–224.

Collier, Paul, and Anke Hoeffler. 2004. “Greed and grievance in civil war.” *Oxford Economic Papers* 56 (4): 563–595.

Dancygier, Rafaela M. 2010. *Immigration and conflict in Europe*. New York, NY: Cambridge University Press.

Enos, Ryan D. 2014. “Causal effect of intergroup contact on exclusionary attitudes.” *Proceedings of the National Academy of Sciences* 111 (10): 3699–3704.

Enos, Ryan D, and Noam Gidron. 2016. “Intergroup behavioral strategies as contextually determined: Experimental evidence from Israel.” *The Journal of Politics* 78 (3): 851–867.

Fariss, Christopher J., and Zachary M. Jones. 2017. “Enhancing Validity in Observational Settings When Replication is Not Possible.” *Political Science Research and Methods* https://doi.org/10.1017/psrm.2017.5.

Fearon, James D, and David D Laitin. 2003. “Ethnicity, insurgency, and civil war.” *American Political Science Review* 97 (01): 75–90.

Greenhill, Kelly M, and Ben Oppenheim. 2017. “Rumor has it: The adoption of unverified information in conflict zones.” *International Studies Quarterly* 61 (3): 660–676.

Habyarimana, James, Macartan Humphreys, Daniel N Posner, and Jeremy M Weinstein. 2007. “Why does ethnic diversity undermine public goods provision?” *American Political Science Review* 101 (04): 709–725.

Hartzell, Caroline, and Matthew Hoddie. 2003. “Institutionalizing Peace: Power Sharing and Post-Civil War Conflict Management.” *American Journal of Political Science* 47 (2): 318–332.

Jha, Saumitra. 2013. "Trade, institutions, and ethnic tolerance: Evidence from South Asia." *American Political Science Review* 107 (04): 806–832.

Kasara, Kimuli. 2016. "Does Local Ethnic Segregation Lead to Violence?: Evidence from Kenya." *Quarterly Journal of Political Science* (forthcoming).

King, Gary. 2013. *A solution to the ecological inference problem: Reconstructing individual behavior from aggregate data*. Princeton, NJ: Princeton University Press.

King, Gary, Martin A Tanner, and Ori Rosen. 2004. *Ecological inference: New methodological strategies*. New York, NY: Cambridge University Press.

Lipsky, Michael. 2010. *Street-level bureaucracy: dilemmas of the individual in public service*. New York, NY: Russell Sage Foundation.

Posner, Daniel N. 2005. *Institutions and ethnic politics in Africa*. New York, NY: Cambridge University Press.

Sands, Melissa L. 2017. "Exposure to inequality affects support for redistribution." *Proceedings of the National Academy of Sciences* p. 201615010.

Shadish, William R. 2010. "Campbell and Rubin: A Primer and Comparison of Their Approaches to Causal Inference in Field Settings." *Psychological Methods* 12 (1): 3–17.

Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2001. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Wadsworth Publishing.

Zachary, Paul, and William Spaniel. 2018. "Getting a Hand By Cutting Them Off: How Uncertainty over Political Corruption Affects Violence." *British Journal of Political Science* .

2 Who Protests? Discretionary Arrests Increase Participation in Collective Action

Abstract

Who protests? Prior scholarship on collective action identifies numerous individual-level causal mechanisms but measures these factors at the group-level. This is because individual-level data on participation or context are difficult to obtain. We address this inferential problem with data from the 2015 Freddie Gray protests in Baltimore, MD. We argue that arbitrary or capricious encounters with civil servants generate grievances. To assess the relationship between this grievance generating mechanism and the decision to protest, we obtain every tweet made during the protests and train a classifier model to estimate who protested. Next, we use a novel algorithm on users' tweet history to estimate where in the city users socialize. We estimate variation in anti-police grievances using these locations and geospatial arrest data from the Baltimore Police Department. Our results show that grievances — measured by exposure to policing events — are significantly correlated with an individual's participation in the protest.

2.1 Introduction

Some of the most memorable and consequential political events stem from protests. By aggregating disparate voices under a single banner, these collective action events have the potential to communicate the preferences of groups that political elites and lawmakers might not otherwise hear (Gause 2017; Davenport 2005; Williamson, Skocpol, and Coggin 2011; Chong 2014). Beyond their expressive value, collective action can change public opinion and influence policy (Madestam et al. 2013; Lohmann 1995). Despite their potential for effecting real policy change, no protest is ever universally attended. This is especially so at the start. For every person who joins in collective action events, others abstain and remain at home. These contextual variations beg the question: who protests?

Prior research on collective action gives unsatisfactory answers to this question. Rather, it offers important insights into related but different research questions such as the determinants of protest onset.[1] Existing formal models show that "tipping points" (Kuran 1989) or "informational cascades" (Lohmann 1994) can spur protests on. Protests begin when the risk of joining is less than citizens' costs of suppressing their anti-government feelings (Kuran 1989: 12-13). Participation then grows when attendance reveals information about social discontent to more moderate citizens (Lohmann 1994: 52-54).

How are these cascades triggered? These aforementioned formal models assume activists form a subset of the population, but do not explain why some groups' preferences are more extreme. Political context can affect behavior by generating grievances (Enos 2014). Scholars of relative depravation such as Gurr (1970) posit that perceived economic inequalities generate grievances in society. These horizontal inequalities among groups motivate participation in collective action in order to secure tangible economic or political benefits (Cederman, Wimmer, and Min 2010; Cederman, Weidmann, and Gleditsch 2011). As the grievances produced by

[1] There is also a well-developed literature on how leaders organize protests. See Walls (2015), Han (2014) and Ganz (2009).

inequality grow in intensity, elites aggregate demands into social movements. This suggests groups at the tails of the distribution of economic goods — either poor or rich — should be more likely to protest.

Despite its intuitive appeal, prior research finds little empirical support in the quantitative literature for the contention that grievances lead to collective action (Fearon and Laitin 2003). As qualitative accounts of collective action continue to stress the role of grievances in motivating participation, this missing empirical link is surprising (Wood 2003). One possible explanation is we measure grievances with "inappropriate conceptualization and imperfect measurements" Cederman, Weidmann, and Gleditsch (2011: 478). While scholars commonly measure grievances at the group-level, relative depravation theories suggest people derive their perceptions from their local context. Rather than rule out grievance as a motivation for participation in collective action, prior studies lack sufficiently granular data to test how perceptions affect behavior.

Though group-level factors influence a given groups' likelihood of participating in collective action, we agree that they are insufficient to explain individual variation. We argue that encounters with state bureaucracy and civil servants can generate individual-level grievances, and that these grievances make it more likely an individual participates in collective action. This is because civil servants, such as teachers, social workers, and law enforcement, provide important information about an individual's place in society. Civil servants, commonly referred to as "street-level bureaucrats," have lax oversight and wide discretion in their decision making (Lipsky 2010). This makes their decision-making capricious. Depending upon an individual's influence, street-level bureaucrats can use their discretion to facilitate or hinder their access to government resources. As few people have sufficient connections to influence bureaucrats, every encounter signals to citizens that the government is at best ambivalent or at worst abusive to their interests (Carrington 2005). This implies the more a citizen interacts with street-level bureaucrats, the more intense their grievances become.[2]

[2] For them to effect, these grievances must be substantively related to the protest. Our theory does not predict that encounters with civil servants drives participation in all forms collective action.

Substantial empirical evidence bolsters this last claim. Encounters with street-level bureaucrats decrease people's trust in government and perceived belonging in their communities (Lerman and Weaver 2014; Uggen and Manza 2002). In this paper, we contribute to the growing literature on the consequences for encountering the "carceral state." Over and above other street-level bureaucrats, police encounters have an especially strong and negative effect on trust in government and political engagement (Kang and Dawes 2017; Tyler, Fagan, and Geller 2014; Wildeman 2014; Uggen and Manza 2002).[3] This negative effect is not isolated to those directly interacting with law enforcement; it can spread through social networks (Burch 2013). Interactions with law enforcement also generates strong grievances due to perceptions of unfair targeting and second-class citizenship (Lerman and Weaver 2014). This suggests that people in high police activity ares should be more likely to engage in collective action related to reforming the police. As grievances are associated with particular institutions, this theory does not suggest increased participation in events that are substantively unrelated.

Grievances alone are an insufficient explanation of protest participation. In all but the rarest of cases, protests are planned and scheduled in advance (Han 2014; Payne 2007; Ganz 2009). Without knowing when and where a protest is scheduled, participation is impossible. We argue that increased demand for information is the mechanism through which grievances translate into participation. Our theory predicts community members with more intense grievances should have greater demand for information for logistical information from community organizers when grievance generating events take place (Zuern 2011; Norris, Walgrave, and Van Aelst 2005; Wood 2003; Murdie and Bhasin 2011).[4] As we do not expect people with low grievances to seek out information about logistics, they should be unlikely to appear at protests.

In most cases, social scientists lack individual-level data on participation and/or grievances. This forces quantitative scholars to use proxies for grievances such as access to foreign media or

[3] This decrease in trust occurs even when an interaction does not result in arrest.

[4] Although it is an important question, understanding activists' choice of when and where to hold protests is beyond this paper's scope.

measures of economic inequality (Kern 2011; Steinert-Threlkeld et al. 2015; Steinert-Threlkeld 2017).[5] These measures, however, are either time-invariant or coarse aggregates. Without granular data, we cannot rule out grievance and perceptions as motivators for collective action.

To address this inferential problem, we introduce a new method that enables us to estimate individual-level social context via social media data. We study the protests that took place in Baltimore, MD in April 2015 against police brutality.[6] One issue when studying the police in the United States is the high correlation between a neighborhood's racial composition and police activity; police maintain a stronger presence and are more likely to use coercive authority in communities of color (Smith 1986). While parsing out the effect of race from police contact would be difficult in these localities, Baltimore exhibits uniquely significant variation in levels of police activity and race.[7] This makes it an ideal test case for our theory.

Our research design is as follows. First, we first obtain every geotagged tweet made within the city in April 2015 and hand code a random subsample according to whether it indicates the person attended the Freddie Gray protests. Second, we use this training set to train a machine learning model to classify the remainder of our corpus of tweets. Third, we purchase the entire tweet history from accounts from our initial sample and then pass every user's tweets through a novel community detection algorithm. This new algorithm, based on earlier work by Rossmo (1999), uses the location where every tweet was posted to estimate where each user spends time throughout their day (Chen, Fariss, and Zachary 2017). These areas, which we refer to as social activity hubs (SAHs), allow us to estimate each user's political context. Finally, we use georeferenced arrest data from 2012-2015 that we obtained from the Baltimore Police Department through a public records request. Using these data, we find a robust and positive association

[5] Other scholars such as Lawrence (2016) and Aytaç, Schiumerini, and Stokes (2017) survey protest participants to overcome this data limitation. Inferences from surveys present a number of methodological concerns such as desirability bias and homophily. We are unable to locate an example of a study that surveys both protest participants and abstainers, making it impossible to empirically assess the differences between these two groups.

[6] As they took place after the death of Freddie Gray in police custody, these protests are commonly referred to as the Freddie Gray protests.

[7] Appendix Figures 2.B.4 and 2.B.5 plot histograms of this relationship. They show there are both high and low police contact zones in predominantly African-American neighborhoods.

between police contact and protest participation.

This paper makes several contributions to the theoretical literature on collective action. First, we show that qualitative accounts were right not to abandon their focus on grievances as motivators for participation. Our results, however, broaden prior studies' conceptualization of grievances to include political context. While material deprivation can produce grievances, our results suggest that the way the state interacts with its citizens affects their behavior. Second, we begin to investigate the microfoundations of compliance and enforcement. As prior studies focus on compliance among institutions, courts, and countries, we know relatively little about the mechanisms through which governments engender compliance with the law among their citizens (Carrubba 2005). Our results suggest additional avenues for research about the downstream consequences of enforcement on political behavior. Third, our results provide clarity about the social mechanisms that influence an individual to choose to join protests even when the costs are high — such as in autocratic societies.

Finally, we introduce a new estimator of social context from social media data. While social media is an increasingly common sampling frame, there was no way to measure where users reside in observational studies. This limits both the types and the quality of inferences that could be drawn from social media. Our estimator, which measures social context down to the neighborhood level, gives researchers an important new tool.

2.2 Who Protests?

Understanding collective action is one of the oldest research questions in social science (Gurr 1970; Olson 1965; Ostrom 1990; Popkin 1979; Tullock 1971). Prior research can be divided into two separate research questions. The first focuses on the link between economic depravation and collective action (Gurr 1970). They posit collective action is more likely in groups within the tails of the distribution of social goods. The second explores how opportunity structures affect participation (Earl et al. 2004; Tullock 1971; Schussman and Soule 2005; Lichbach 1994). In this section, we describe these prior studies' insights before noting there is no satisfactory explanation of individual-level variation in participation *within* groups. To address this gap, we propose a novel theory of how social context affects grievances, which in turn help explain who protests and when.

2.2.1 Relative Depravation and Inequality

While they highlight different causal mechanisms, relative depravation-based and inequality-based theories tie collective action to grievances produced by an individual's access to material goods (Gurr 1970, 1993; Cederman, Weidmann, and Gleditsch 2011). For relative depravation scholars, the primary source of social discontent is unfulfilled material aspirations. The first step in the causal process is for an individual to perceive a gap between the standard of living they believe they deserve and what they actually have. Taking stock of this gap creates grievances, which inter-personal and inter-group comparisons then exacerbate. As this perceived gap grows, moreover, it produces more intense grievances. As Regan and Norton (2005: 320-321) summarize, the "causal mechanism is the psychological process in which an individual compares their current situation against their expected standard of living." These perceptions translate into collective action once elites aggregate channels these feelings into a social movement.

Although both focused on standard of living, a different causal mechanism generates grievances from inequality. If relative depravation focuses on perceived inequalities, then inequality highlights real ones. Grievances begin once an individual compares their level of material wealth to others in society. These grievances grow more intense as the gap in material wealth grows between the individual and their reference group. Inequality does not exclusively generate grievances among the poor. As Cederman, Weidmann, and Gleditsch (2011) and Boix (2008) argue, economic inequality can also incentivize the rich to mobilize to prevent economic transfers to the poor. Economic inequalities translate into collective action after elites aggregate demands for social and economic justice into a social movement.

While theories of relative depravation and inequality generally link the distribution of material wealth to grievances, wealth is not necessarily the only such source. Grievances come from unequal access to whatever the relevant economic cleavage in society is. Prior scholarship identifies sources of grievances as diverse as access to land rents (Thomson 2016; Acemoglu and Robinson 2006); misallocation of investment (Robinson and Torvik 2005); and natural resource rents (Ross 2004). Moreover, the inequality need not be economic. Perceived status inequalities can also trigger collective action as groups use protests to reposition themselves in the social hierarchy or demand more rights (Horowitz 1985; Chong 2014; Willer 2009).

Regardless of the precise source of the inequality, these theories observable implication is that collective action is more likely in communities along economic extremes than those closer to the median.

2.2.2 Opportunity Structures

Even if relative depravation and inequality are intuitively appealing, there is very mixed empirical support that links the distribution of material goods to collective action (Fearon and Laitin 2003; Hegre and Sambanis 2006; Lichbach 1989). Moreover, relative depravation loses some of its theoretical appeal under closer inspection. As Snyder and Tilly (1972) note, every society is replete with aggrieved, angry, and frustrated people. Especially in poor communities, real and perceived inequality and depravation is near-universal. Yet, protest and other forms of collective action remain rare.

If material depravation is universal and time-invariant, it is insufficient to explain participation in collective action. In their most reduced form, theories of political opportunity structure suggest collective action when "the political environment that provide[s] incentives for people to undertake collective action" changes such that it affects potential participants' "expectations for success and failure" (Tarrow 1994: 85). Such changes can include exogenous developments such as technological development (Steinert-Threlkeld 2017) or endogenous weakening of state repressive institutions (Fearon and Laitin 2003). In all cases, collective action is determined by citizens ability to organize and the states capacity to manage opposition.

The first set of political determinants of contentious politics are factors that influence citizens willingness to organize protest and realize potential risks from collective action. Citizens considering joining in protests encounter challenges from both the state and others in their group (Kuran 1989; Lohmann 1994). The state can use violence and repression to diminish turnout (Davenport 2010; Davenport, Soule, and Armstrong 2011; Lorentzen 2014). Protestors confront collective action problems in both the presence or absence of repression (Olson 1965). Even when protests to secure policy redress is Pareto-improving — when the change in allocation over the desired good makes one or more individuals better off without making another worse off — each individual should rationally abstain and free-ride off of others. This suggests that a time-varying mechanism enables citizens to organize and overcome this collective action problem.

What separates first movers from other citizens in their ability to solve collective action problems? The literature identifies two answers to this question. The first is that some citizens — such as activists and community organizers — might receive especially high amounts of utility from political engagement. This increased utility makes activists "extraordinarily willing to take risks for a worthy political cause" (Lawrence 2016: 4). While the literature is unclear about the exact source of this increased utility, it might simply derive from the "pleasure of agency" (Wood 2003: xv-xvi). The other is that first movers are unusually interested in acquiring some selective incentive such as political power or resources (Popkin 1988; Schneider and Teske 1992). Another mechanism that solves the collective action problem for first movers is selective incentives (Lichbach 1994, 1998; Mason 1984).

2.2.3 Grievances and the Selective Incentive for Information

In this section we devise an argument that links information seeking and grievances. Individuals interact with state security agents in a variety of settings. Sometimes these interactions are without incident. Under other circumstances however, grievances are generated by the interaction. If well-known means to address the grievance exist then information seeking takes places through established institutional channels (e.g. through the judicial system). However, if such institutions do not exist or are not trusted then information seeking must, by necessity, take place outside of established institutional channels. This type of information seeking behavior creates a selective incentive to organize by the individuals directly affected by an interaction that generates the grievance. When the aggrieved do not trust the courts or the police, then organizing a group to seek information about the event by protest becomes an important alternative option. These first movers solve the collective action problem for themselves because they have a selective incentive to organize. Beyond this, as these individuals organize into a larger groups the costs of joining in the protest decreases, which therefore solves the collective action problem for others.

The collective action problem exists because the costs of participating in action against the state are high and if the movement is successful all members of the group enjoy the benefits regardless of whether they participated in the movement or not (Lichbach 1994, 1998; Mason 1984; Olson 1965). In order to solve collective action problems selective incentives may be used to compel or even coerce participation (Lichbach 1994, 1998; Mason and Krane 1989). Selective incentives are private or club goods available to individuals participants in the collective action but not every member of the group. Selective incentives can include wages, paid now or in the future, or other incentives obtainable by the individual through participation in the group.[8]

[8] Lichbach (1994) argues that collective actions problems are solvable when selective incentives are used; however, selective incentives applied without reference to a collective goal, often represented by an ideology, do not result in successful movements. Specifically, Lichbach (1994) argues that the collective action problem is more likely to be solved under three conditions: (1) private goods are available as benefits, (2) the costs and benefits applied to each peasant is different given the level of participation, (3) the distribution rule that governs access to the good is contingent on participation. According to Lichbach (1994), collectively supplied benefits are "public goods", which are nontrivial and excludable. "The term thus covers objectives that are both reformist (e.g., having the existing

Building on prior research on collective action, we argue that individuals with more protest-relevant grievances should be more willing to participate in collective action because they have a selective incentive to obtain information — an instrumental motive. Instrumental motivation theories suggest that horizontal inequalities should largely determine behavior (Gurr 1970; Cederman, Weidmann, and Gleditsch 2011). Once activists overcome collective action problems, members of social classes who are deprived relative to the reference group join in. While this helps explain why some classes are — on average — more likely to mobilize than others, it does not explain individual-level variation *within* these groups.

In this section, we develop a novel theory that explains why some join protests while others abstain within the same social grouping. Our theory highlights the role played every citizen's social and political context in producing grievances (Enos 2016; Enos and Gidron 2016; Enos 2014; Sands 2017). Even within the same community, citizens might have widely divergent lived experiences. Some community members might be satisfied with the status quo, while others might be highly motivated to seek political redress.

The literature on conflict onset highlights a number of mechanisms that produce grievances. Horizontal inequalities, or "inequalities in economic, social or political dimensions or cultural status between culturally defined groups," are one such mechanism (Stewart 2008: 3). These inequalities might arise when a group perceives it does not have its fair share of society's wealth (Collier and Hoeffler 2004). Inversely, poor economic conditions can also make groups more likely to mobilize in order to protect the current distribution of goods (Miguel, Satyanath, and Sergenti 2004).

Demands for prestige and status can also mobilize groups into collective action (Sambanis 2001; Cederman, Wimmer, and Min 2010). The literature on ethnicity and civil conflict frequently links nationalism and demands for self-determination with mobilization (Fearon, Kasara, and Laitin 2007; Buhaug, Cederman, and Rød 2008). Laws and institutions that exclude specific

government supply a new road) and revolutionary (e.g., taking over the government and supplying your own road" (Lichbach1994: 388, note 25).

groups from power serve as a focal point for organizing and solving collective action problems. This dynamic can be seen in the United States from the Civil Rights-era protests as well as the ongoing LGBT equality movement (Chong 2014; Davenport 2007).

These prior theories intentionally "[shift] the explanatory focus from individualist to group-level accounts" of mobilization (Cederman, Weidmann, and Gleditsch 2011: 477). While group-level accounts help explain why — on average — some groups are more likely to mobilize than others, they cannot explain variation in participation *within* groups. While we concur with theories focusing on horizontal inequalities, we focus instead on how context affects how individuals perceive their position within their cultural group. In our empirical work below, we explore how interactions with bureaucrats affect grievances.

As Lipsky (2010) argues, the quality and nature of an individual's interactions with government employees are one important signal of their context and position within their group. These "street-level bureaucrats," such as social workers, police officers, or school teachers, wield enormous power over citizens' lives. They determine whether a permit is issued; a child is removed from their parent's custody; or a fine is imposed. Importantly, these bureaucrats have tremendous discretion and as such can determine how much access a citizen has to government services. As citizens use these interactions to update upon their support for the government, every interaction with a street-level bureaucrat has the possibility of generating new grievances (Lipsky 2010; Butler and Broockman 2011; White, Nathan, and Faller 2015).

Although our theory generalizes to interactions with all bureaucrats, our empirical focus in this paper is the effect of encounters with police officers on grievances and mobilization. Interacting with police officers is rarely pleasant and people of color must be especially vigilant to protect their physical integrity (Davenport 2005, 2010). This observation is highlighted by the growing literature on the consequences of interactions with the "carceral state," which shows that police-citizen interactions decrease trust in government and participation in politics through formal institutions (Kang and Dawes 2017; Tyler, Fagan, and Geller 2014; Lerman and Weaver

2014; Wildeman 2014; Uggen and Manza 2002). These negative effects also propagate locally within social networks to those who did not interact directly with the police (Burch 2013).

This observation suggests that police encounters produce grievances at the individual-level. over and above those horizontal inequalities generate, these additional grievances from social interactions explain who joins during collective action events.

How do these grievances help individuals overcome collective action problems? We argue that grievances affect demand for information about protests and opportunities for collective action. Logistical information such as the time and date of a planned protest is necessary for participation. Although we do not test this mechanism directly, we expect grievances to increase information-seeking behavior through social media or personal social networks.

Acquiring this information is important because it is inefficient for organizers to try to mobilize people who prefer the status quo. Second, after identifying aggrieved community members, they use these individuals to disseminate information within their social networks. Because organizers cannot talk to everyone in the community, they rely upon the people they contact to spread information through their social networks. This exposes individuals with higher levels of grievance to community organizers and activists (Zuern 2011; Norris, Walgrave, and Van Aelst 2005; Wood 2003; Murdie and Bhasin 2011).

2.3 Research Design

Studying collective action presents numerous measurement problems. While most theories make individual-level predictions about behavior, measures of participation and/or grievances are rarely available at such fine resolution.[9] To surmount this problem, scholars typically rely on aggregate measures such as access to foreign media or self-reports of behavior in surveys

[9] As Cederman, Weidmann, and Gleditsch (2011: 478) note, the “formidable problems of data availability associated with the uneven coverage and comparability of” individual-level measures of grievances and/or mobilization, “most scholars have had to content themselves with selective case studies or statistical samples restricted to particular world regions.”

(Kern 2011; Steinert-Threlkeld et al. 2015; Steinert-Threlkeld 2017; Lawrence 2016; Aytaç, Schiumerini, and Stokes 2017). Each approach limits the types of behavioral inferences that can be made. Inferring individual behavior from aggregate data can produce highly inaccurate, biased estimates (King 1997). Known variously as aggregation bias or the ecological inference problem, this statistical property makes inferences about individual-level behavior derived from aggregate data particularly fragile.

While they do not suffer from aggregation bias, surveying protestors poses different challenges. First, social desirability bias likely influences respondents' answers in ways that are correlated with their social context and the protest's outcome. A respondent might overstate their participation after successful protests or diminish it after failed ones. Second, it is incredibly difficult to validate the representativeness of a sampling frame. This issue is non-trivial because scholarship on homophily suggests that participants will segregate within protests (Clack, Dixon, and Tredoux 2005).

To overcome these challenges, scholars have recently turned to social media data to measure patterns of protest and mobilization (Steinert-Threlkeld et al. 2015; Steinert-Threlkeld 2017; Lawrence 2016). While social media data has the advantage of permitting scholars to measure protestors' social networks and explore how information spreads within networks, it has the drawback of lacking the contextual and demographic information that comes with surveys. Twitter does not disclose its users race, age, or gender and offers only rudimentary tools to infer location.[10] This forces scholars to use aggregate data, subject to the bias described above.

By introducing a novel method to estimate social context from social media data, we propose a novel solution to these inferential problems. We test for how grievances influence protest participation during a highly contentious and violent series of protest and repression events that occurred during April 2015 in Baltimore, MD following the death of Freddie Gray in Baltimore Police Department (BDP) custody. We focus on Baltimore because it fits the criteria of

[10] Researchers have access to each account's time zone and language setting.

a "typical" case, in which "cases are intended to represent descriptive features of a broader set of cases" (Gerring 2008). Typical cases are particularly beneficial for theory testing. The correlation between community demographics and police activity is very strong in the United States (Smith 1986). Unlike other American cities, Baltimore has numerous communities of color with high and low police activity. Without this variation, it would be impossible to parcel the effect of race out from that of police contact.

BPD officers arrested Gray on April 12, 2015 for suspected possession of an "illegal switchblade."[11] Gray fell into a coma for reasons that are disputed while being transported in a police van. He did not recover from his injuries and died on April 19. To denounce Gray's alleged mistreatment and BDP brutality, protestors began gathering in front of the Western District Police Station on April 18. These protests grew rapidly in size throughout the week and eventually reached several thousand people and only ended after the Maryland National Guard restored calm to the city by imposing a mandatory curfew from April 28 to May 3. These protests, collectively referred to as the Freddie Gray protests, are our empirical focus.

Our research design is as follows. First, we purchased every geotagged tweet made from within Baltimore, MD from April 16, 2015 to May 4, 2015, i.e. during the Freddie Gray protests.[12] This generated a total of 111,440 tweets from 7,884 unique users. Second, we hand coded 5,000 randomly selected tweets as to whether each tweet indicated participation in the protests.[13] Third, hand coded tweets serve as a training set for a classifier model that we use to classify our remaining corpus of tweets as to whether they indicate protest participation. Our results suggest that approximately 5% of our corpus of tweets indicate protest participation.

Third, we purchase the entire account history from users in our data. Fourth, after

[11] While officers testified they believed Gray's knife was illegal, the Maryland State's Attorney for Baltimore later clarified that he was in fact in possession of a "spring-assisted knife" that was legal under Maryland law (Blinder and na 2015).

[12] Location sharing is an opt-in feature and is disabled by default. One concern is that users who opt into sharing this data are systematically different than those who do not. Addressing this question, Pavalanathan and Eisenstein (2015) find that users who enable geolocation are demographically similar to other Twitter users.

[13] Two research assistants read and coded the subsample separately. The intercoder reliability was over 99%.

obtaining the account histories, each account is passed through a novel community detection algorithm (Chen, Fariss, and Zachary 2017). Based on earlier work on community detection by Rossmo (1999), our algorithm identifies clusters in locations were users tweet in order to estimate their movement patterns. These areas, which we call social activity hubs (SAHs), enable us to estimate users' political context as measured by where they spend time. Finally, we measure grievances with georeferenced arrest data we obtained from from the Baltimore Police Department through a public records request. Using these data, we find a robust and positive association between police contact and protest participation.

2.3.1 Protester Classification

In order to develop a classifier model of protest-related tweets, we manually coded a random subsample of 5,000 tweets drawn from our full corpus. We assign tweets a 1 when it indicates the user was physically present at the Freddie Gray protests and assign tweets a 0 when they are unrelated to the protests.[14] For examples of both protest and non-protest related tweets, see Figure 2.1:

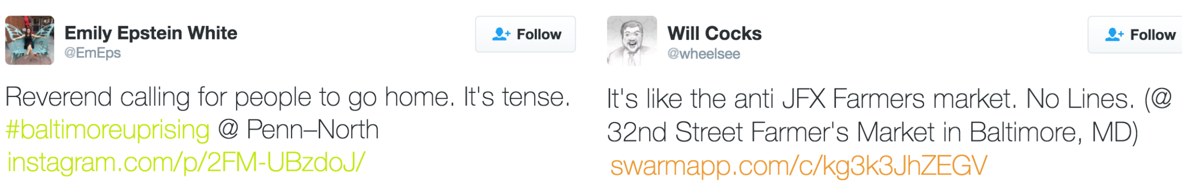

Figure 2.1: **Left:** A protest-related tweet referencing ongoing discussions about whether protestors would comply with the curfew order. **Right:** A non-protest related tweet.

Classification

We estimate the probability the tweets outside of our training set indicate the poster attended the Freddie Gray protests in person. After preprocessing, we randomly divide our

[14]624 tweets, or approximately 12%, of our training set suggested the user attended the protests.

manually coded tweets into a training (80%) and a test (20%) set for cross-validation.[15] We fit two models after performing a grid search: regularized logistic regression (LR) and a support vector classifier (SVC) with feature selection.[16]

To estimate the probability each tweet belongs to a protestor, LR models fit a linear regression using all features as regressors. The output is then the estimate of the probability with which each observation belongs to either category using the logistic function. SVCs identify the minimal support vectors, which distinguish texts in one class versus another. Specifically, we collapse our training data into a matrix (**X**), where each item x_{ij} indicates the j features used by Twitter user i. As we coded whether each Twitter user protested (p_i,) we can write our training data as (x_{ij}, p_i). The SVC attempts to separate this data such that we can describe the best hyperplane whereby one set of Twitter users has $p_i = 1$ while the other has $p_i = 0$..[17]

SVCs identify a vector (w) such that w characterizes the two hyperplanes that maximize the separation between the two classes, i.e. the w whereby $w \cdot x_i - b \geq 1$ is true. In other words, SVCs minimize $\|w\|$ such that $\forall_i = 1, \ldots, np_i(w \cdot x_i - b) \geq 1$. The SVC classifies the remaining corpus of tweets by projecting them onto the hyperplane generated previously. This projection is then converted into a probability distribution after fitting a sigmoid (Platt 1999).

As our SVC and LR both performed well, we averaged their predictions for our remaining corpus using soft voting, which sums the predicted probabilities across an ensemble of classifiers. We then evaluate the quality of our predictions with their Matthews correlation coefficient.[18] After crossvalidation on our test set, we estimate that the final accuracy of our model using soft

[15] For a more detailed discussion of our preprocessing strategy, please see Appendix Section 2.A.1 in our online appendix.

[16] Although naive Bayes is a common text classification algorithm, our initial grid research suggested its accuracy for our particular application was significantly lower than other approaches.

[17] A hyperplane is the set of points in vector x such that $w \cdot x - b = 0$.

[18] Matthews correlation coefficients (MCC) are a correlation coefficient between the observed and predicted binary classification (Matthews 1975). As a measure of the quality of binary classifications, MCCs are particularly robust to instances when the two classes are of very different sizes. MCCs can be calculated from the confusion matrix as:$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN(FN+FP)(TN+FN)}}$, where TP is the number of true positives; TN the number of true negatives; FP the number of false positives; and FN the number of false negatives.

voting to be 95.12% (σ = 0.50) with a false positive rate of 13.92% (σ = 3.53).

In order to maximize the size of each user's corpus of tweets, we aggregate our final estimation up to the account-level. We then fit our final model onto the entire uncoded set. We code a user as attending the Freddie Gray protest if at least one of their tweets is protest-related. At the user level, we simulating classification errors to estimate our false positive rate: 8.21% (σ = 1.80). Figure 2.2 shows the proportion of users and tweets that our classification procedure estimates indicate protest participation:

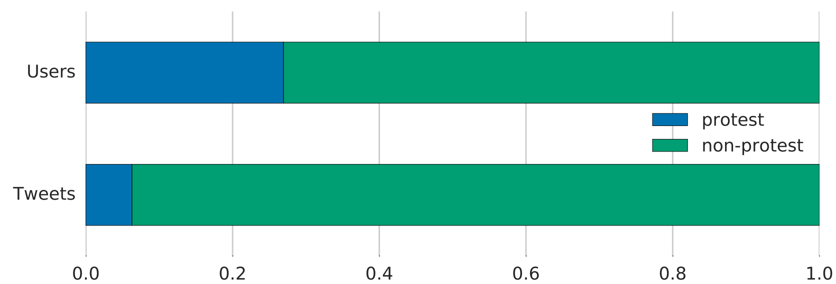

Figure 2.2: This bar plot shows the proportion of tweets and users that indicate protest participation. The proportion of users classified as participating in the protest exceeds the proportion of tweets discussing the protest because protestors used Twitter to discuss both the protest as well as other subjects.

2.3.2 Social Activity Hubs

Although the above enables us to identify which users attended the Freddie Gray protests, it does not provide us with information about the users' social and political context. We argue that context conditions political behaviors such as protest participation. More explicitly, this implies that exposure to social and political cues present within the environments affects political behavior. Testing this argument requires some method to place users in geographic space. In this section, we describe our novel method that estimates Twitter users' "social activity hubs" (SAHs), or the centroid of the areas where users spend time throughout the day (Chen, Fariss, and Zachary 2017).

Prior to estimating users' SAHs, we collected each user's historical tweets using Twitter's API, which we called through the TwitteR package in R (Gentry 2015).[19] Each geotagged tweet is treated as an observed incident of the user's movement patterns, and from the collection of all observed incidents, we estimate the user's SAHs. As the availability of information associated with each Twitter account differs, our SAH model, summarized in algorithm 3, is conditional on what this information affords, defaulting to more basic models where data availability is low. More specifically, we define SAHs in two ways described in more detail below, based on posterior quantities of a Dirichlet process mixture (DPM) model for spatial data (Verity et al. 2014). Where these estimates are unavailable due to convergence failure in the MCMC (implemented in the Rgeoprofile package for R, Stevenson et al. 2014), we take the incident closest in Euclidean distance to the spatial mean of all observed incidents as the SAH. For users with only one observed incident ($n = 197$), that incident is taken as the SAH.

In the remainder of this section, we describe the DPM model in more detail, discuss how we use its posterior quantities in our SAH model, and explain how we account for uncertainty in the model. We conclude by presenting some ethical considerations.

Dirichlet Process Mixture Model for Spatial Data

For user whose tweets contain enough information regarding their movement patterns, we use the Dirichlet process mixture (DPM) model of geographic profiling as the basis of our SAH model. DPM models for spatial data, based on prior geographic profiling models in criminology (O'Leary 2010; Rossmo 1999), was first described in (Verity et al. 2014) where it was applied to spatial epidemiology. More recently the model was applied in an attempt to determine the identity of graffiti artist Banksy (Hauge et al. 2016). The intuition of the DPM model for spatial data is to sort a set of observed incidents in physical space into clusters originating from different

[19] Twitter only allows access to users' 3,200 most recent tweets through its API. Tweets were collected between July 19, 2016 and August 27, 2016. API Documentation available here: `https://dev.twitter.com/rest/reference/get/statuses/user_timeline`.

source locations without prior assumptions about the number of clusters that exist. For our present purposes, the DPM model is preferred over alternatives that require a fixed number of clusters (including those with a single cluster) because individuals are likely to vary in terms of their movement patterns for we have no prior information. Where there are multiple clusters, especially when they are highly dispersed, a misspecified number will result in inaccurate source location estimates skewed by "outliers" which are actually observations originating from another source. The DPM model rectifies this by estimating the number of sources based on the observed data. The flexibility afforded by this feature is especially desirable given the large number of Twitter accounts we are working with, as it is not feasible to tweak the SAH model for each Twitter account individually. A DPM model is not without assumptions, which are apparent from the technical description below. In short, by employing the DPM model, we are assuming that individuals can have multiple SAHs from where their movement outward follows identical distributions, which in this implementation we specify as a bivariate normal distribution with standard deviation of approximately two miles.[20]

More explicitly, the DPM model we use, adapted for spatial data by Verity et al. (2014), is as follows. For each Twitter user, define a two-dimensional sample space with a finite grid of cells as Ω, in which each cell $\boldsymbol{\omega} = (\omega^{(1)}, \omega^{(2)})$ is a vector containing the latitude and longitude in decimal degrees of a geocoordinate. The set of n geocoordinates obtained from geotagged tweets $\mathbf{x} = \mathbf{x}_{1,\ldots,n}$ is assumed to be the result of independent draws from a mixture of a countably infinite set of bivariate normal distributions centered on $\mathbf{z} = \mathbf{z}_{1,\ldots,\infty}$, each with a variance of σ^2 ($\sigma = 0.05$). Both $\mathbf{x}$ and $\mathbf{z}$ are defined on Ω. The prior distribution of the set of $\mathbf{z}$ is assumed to be a bivariate normal centered on the mean of $\mathbf{x}$, with a variance of τ^2 (τ is set to the largest distance in either longitude or latitude). c_i is a categorical variable that assigns $\mathbf{x}_i$ to source $\mathbf{z}_{c_i}$, and is drawn from a Dirichlet process, specifically the Chinese Restaurant Process which has a concentration parameter α drawn from a diffuse hyper-prior (specifically $h(\alpha) = ((1+\alpha)^2)^{-1}$)

[20] Specifically 0.05 decimal degrees which translates to approximately 2 miles in Baltimore, MD.

and a base distribution that is the bivariate normal (with mean $\mathbf{x}/n$) discussed above. The above is formally represented as,

$$\begin{aligned}
\mathbf{x}_i|\mathbf{z}_{c_i} &\sim \mathcal{N}(\mathbf{z}_{c_i}, \Sigma = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}) \\
\mathbf{z}_{1,\ldots,\infty} &\sim \mathcal{N}(\mathbf{x}/n, \mathrm{T} = \begin{bmatrix} \tau^2 & 0 \\ 0 & \tau^2 \end{bmatrix}) \\
c_i &\sim CRP(\alpha) \\
\alpha &\sim \mathcal{H}
\end{aligned} \tag{2.1}$$

Exact computation of posterior quantities are intractable when the number of observations is high ($n > 10$ being a useful rule of thumb; see Verity et al. 2014: for analytical solutions to relevant posterior quantities), but can be estimated using MCMC methods (Neal 2000; Verity et al. 2014), which is implemented in the R package Rgeoprofile 1.2 (Stevenson et al. 2014). The MCMC algorithm (`RunMCMC()` presented in algorithm 4) is based on a Gibbs sampler that alternates between draws of of source locations $\mathbf{z}_{c_i}$ and cluster assignment c_i for all $i = 1, \ldots, n$ observations. The algorithm returns, for each $\mathbf{x}_i$, its cluster c_i; and for each unique cluster c_j, its spatial mean $\mathbf{z}_j$.

Local Minima and Cluster Mean Submodels

As introduced earlier, we use the posterior quantities obtained from the DPM model in our SAH model in two ways. For the *local minima* submodel, begin by defining $\mathbf{S} \subseteq \Omega$ as the grid bound by the minimum and maximum values of the set of observed $\mathbf{x}$. Next, for every cell $\mathbf{s} \in \mathbf{S}$, rank $\mathbf{s}$ according to the sum of its distances to each source location $\mathbf{z}_j$ over all posterior draws, where distance is not linear but weighed by the inverse of the bivariate normal density around $\mathbf{z}_j$. In concordance with existing geoprofiling approaches (e.g. Rossmo 1999), ranks are transformed to hit scores on $[0, 1)$, but remain functionally equivalent in that lower is better and all values are distinct.[21] The type of hit score surface is traditionally used as a surface for search priority of source locations (Rossmo 1999; Verity et al. 2014). On this surface, we find all m local minima (i.e. locations with higher priority) within an approximately two mile radius (0.05 decimal degrees) and define a user's SAHs as the set of m observed $\mathbf{x}_i$ closest to these local minima. For the *cluster mean* submodel, we define a user's SAHs as the set of observed $\mathbf{x}_i$ closest to the set of estimated source locations $\mathbf{z}$ averaged across posterior draws.

Based on each user's geoprofile, we take the most likely sources, up to three locations, and set the incidents (i.e. geocoded tweets) closest to them as the user's SAH, and the census tract associated with that location as the user's SAH neighborhood. Not all estimated SAHs are within Baltimore City bounds, but given the constraint that we only have arrest data from the Baltimore Police Department, described later, we dropped all SAHs not in Baltimore. Results from the DPM model for all users are presented in Figure 4.1. For users with multiple estimated SAHs within Baltimore, we take the average of the census tract characteristics across all estimated neighborhoods, but note that given the parameters of the DPM model (i.e. the standard deviation of the bivariate normal distribution in the migration profile), it is rare for users to have more than two estimated SAHs within Baltimore.

Figure 4.1 illustrates the SAHs estimated under both models in relation to the hit score surface produced by the DPM model.

Uncertainty in DPM Model

In earlier applications of the DPM model to spatial data, there is justifiably less of a concern over the uncertainty of estimates. However uncertain, the expected values of $\mathbf{z}$ is what informs a search that must take place. Existing implementation of the model (Stevenson et al. 2014) therefore do not readily yield uncertainty measures. For inferential modeling, however, measures of uncertainty feature much more prominently. In order to account for uncertainty in

[21]The two computational steps above are implemented in the `ThinandAnalyse()` function in Rgeoprofile 1.2 (Stevenson et al. 2014).

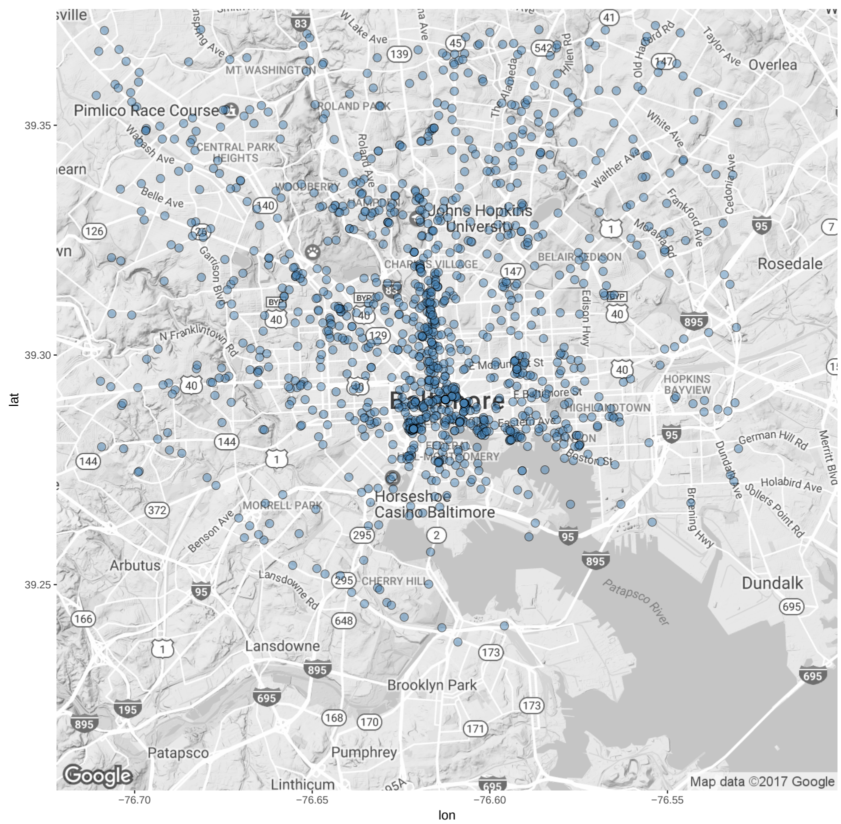

Figure 2.3: Geographical distribution of social activity hubs estimated using the DPM model.

our SAH model, we take 2000 draws from the posterior distribution of the DPM model, thinned to 100 samples, and use this information to determine a set of corresponding SAHs following both the local minima and cluster mean submodels. Specifically, for the local minima submodel, instead of computing a hit score surface based on all posterior draws, we do so for each draw independently; and for the cluster mean submodel, $\mathbf{z}$ is not averaged across posterior draws. SAH estimates are stored for each posterior draw, forming a posterior distribution of SAHs. This distribution can be used in subsequent statistical modeling to account for uncertainty associated with the SAH model. In remainder of this section, we illustrate using a specific Twitter account what the uncertainty about the two DPM-based SAH submodels look like. This particular account was chosen because it is illustrative. The level of uncertainty associated with this account, based on visual inspection, is neither particularly high nor low.

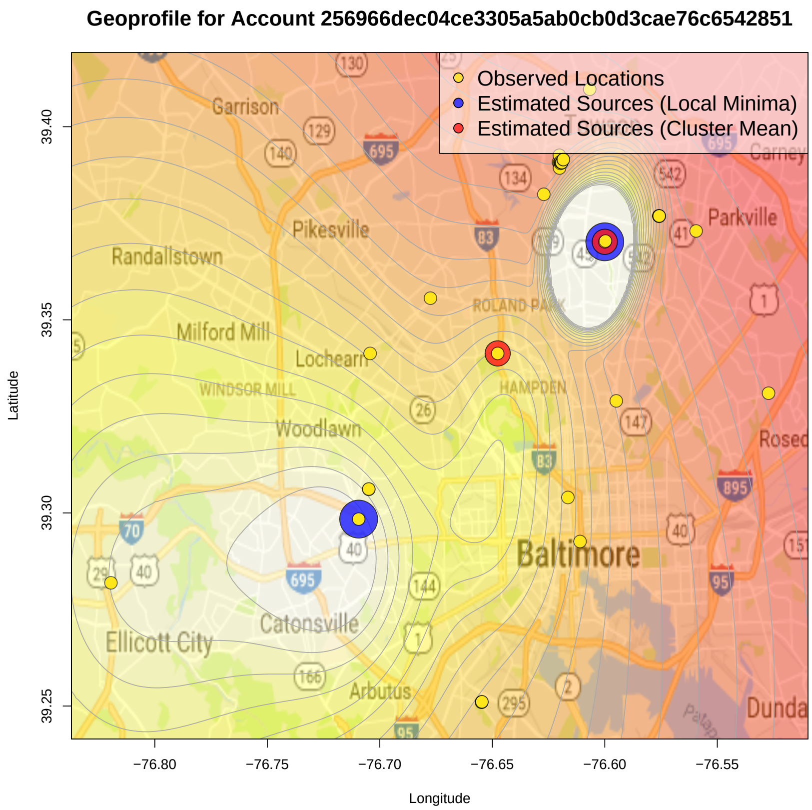

Figure 2.4: Example of a user's estimated social activity hub locations in relation to the hit score surface produced by the DPM model. Yellow points are observed incidents. Points in blue are determined by the local minima submodel. Points in red are determined by the cluster means submodel.

In order to visualize the level of uncertainty about our SAH estimates, we plot the variation in hit scores associated with each potential source across all posterior draws. Future work might benefit from a formal quantification of the type of uncertainty discussed here. Specifically, in Figure 4.2 (which corresponds to the hit score surface in Figure 4.1), each horizontal line documents the hit score of a particular source location as it varies across posterior draws. The highlighted lines are the sources chosen as the SAHs from the combined posterior draws, which may differ depending on the model used. Variation in the hit score indicates changes in the topography of the hit score map across posterior draws. This does not, however, necessarily mean there is uncertainty about the SAH estimates, which arise when changes in the topography are large enough to induce changes in the hit score rank of potential sources relative to each other, indicated by crossing lines.

In this particular case, both methods led to the potential source location with the lowest hitscore as one of the two sources. In this particular example, we see that one source (the one highlighted in blue) has very little uncertainty about it, while the other (in red) has higher uncertainty. The blue line is not straight, meaning that the hitscore map topography around it changes from draw to draw, but it remains consistently low and never crosses with another line, meaning that it is always the source with the best hitscore across all draws.

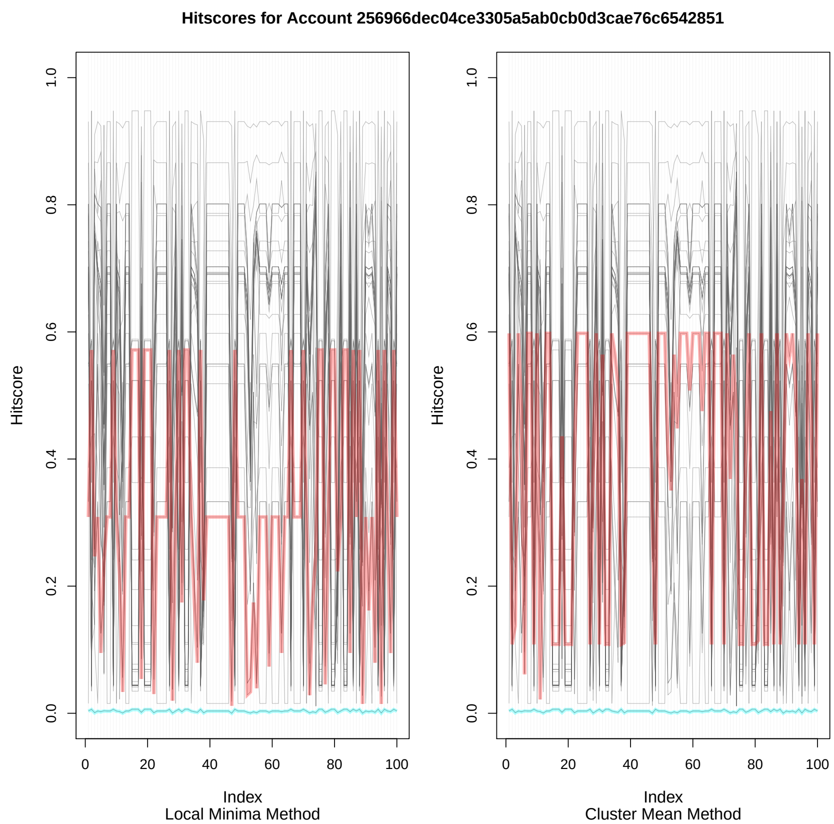

Figure 2.5: Example of a user's estimated social activity hub locations in relation to the geoprofile produced by the DPM model. Blue points are observed incidents. Red points indicate social activity locations under the DPM model and blue ones indicate points from the SAH neighborhood under the modal-tweet model.

Ethical Considerations

As our research strategy enables us to estimate Twitter users' social activity hubs, we take several steps to maintain anonymity and protect users from potential harm. First, our sample is only taken from users who had opted into sharing their location with Twitter. By default, Twitter does not record the location where a tweet was posted. Instead, users must change their phone's settings to give Twitter permission to record their location via GPS. Second, we anonymize Twitter account names by applying a cryptographical hash. Hashing account names prevents anyone with access to either the original or replication data to identify and/or locate the users in our study.

2.3.3 Grievance Measurement

The Freddie Gray protests were directed against the Baltimore Police Department (BPD), which was perceived as hostile to the city's African-American community in the wake of widely publicized cases of police brutality. We proxy for the concentration of anti-BPD grievances using the location of arrests made by the BPD.

These data were obtained through a long-running public records request to the BPD. The data, released to us in 2017, contains records from the years 2012-2015 inclusive. In our initial request, we asked for copies of every arrest's narrative and/or arrest super form in order to code the amount of force used in each arrest. As these reports are not digitized by the BPD, releasing them would require years of redactions and organizing. As such, our request to access these reports was denied by the Chief of the BPD's Legal Affairs Division as "unduly burdensome."

Instead, the BPD provided arrest logs for all felony arrests. These logs contain the date the arrest was made; the location where the arrest was made; and the charges. Any personal information about the arrestee was redacted by the BPD under various subsections of the Maryland Public Information Act related to the release of personally-identifiable information.

We then convert the street address provided with each arrest to latitude and longitude via Google Map's public API. We use data from the 34,880 felony arrests made by the Baltimore Police Department from January 1, 2012 to April 17, 2015. Figure 2.6 shows the location of all unique visits by the BPD that resulted in at least one arrest between 2012 and 2014.

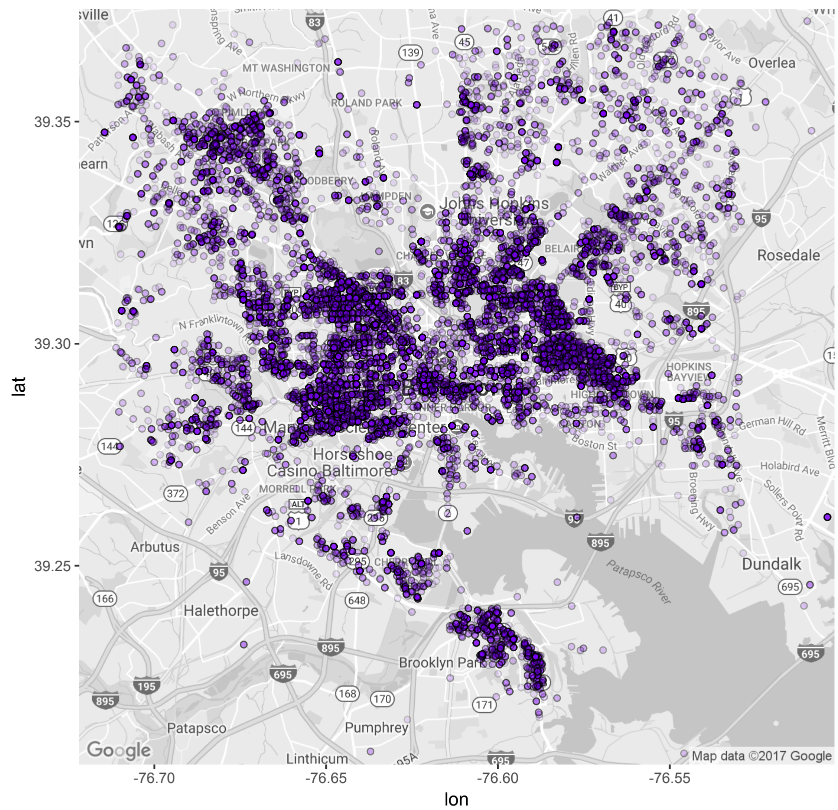

Figure 2.6: Geographical distribution of all unique visits resulting in one or more arrests by the BPD in 2012 and 2015. A small jitter (0.0005 to 0.001 degree decimals) is added to the coordinates.

We use the arrest data to construct two measures of anti-BPD grievances, which is a result of exposure to policing. The first measure, *Total Arrests*, is simply the number of arrests within a census tract. The second measure, *Visits*, treats all arrests at a given location at the same time as a single incident within the census tract. Figure 2.7 shows the density of both measures of exposure to policing in Baltimore census tracts.

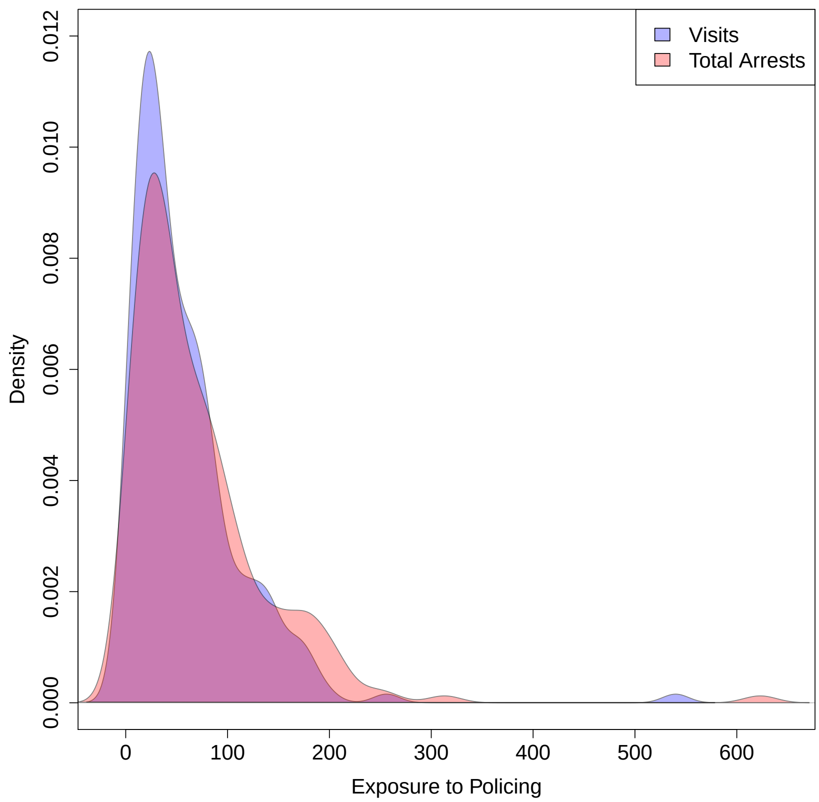

Figure 2.7: Density of both measures of exposure to policing in Baltimore City census tracts.

2.3.4 Control Variables

In order to control for characteristics of each user's home neighborhood, we access information about the racial and demographic makeup of each Baltimore census tract as provided by the U.S. Census Bureau. Specifically, we include in our model the proportion of the population that is (1) Black, (2) Hispanic, (3) female, (4) between the ages of 15 and 24 (i.e. the youth bulge), (5) at least a high school graduate; (5) the median income, the (6) poverty and the (7) unemployment rates; and (8) the overall population. The distributions for these control variables are presented in Appendix Figure 2.B.3.

Our final model is as follows, with two types of exposure to policing as described above.

$$\begin{aligned}
Protest \sim\ & Exposure\ \ to\ \ Policing + Proportion\ \ Black + Proportion\ \ Hispanic \\
& + Proportion\ \ Female + Youth\ \ Bulge + Median\ \ Income + Poverty\ \ Rate \\
& + Unemployment\ \ Rate + Proportion\ \ Highschool\ \ Graduate \\
& + Population + Population^2 + \varepsilon
\end{aligned} \tag{2.2}$$

2.4 Results

Results from our models are presented in Table 2.1 and Table 2.D.1. The substantive effect of exposure to policing is graphically presented in Figure 2.8.

To understand the substantive impact of an additional year of tenure on homicide, we estimate the marginal effect of an additional year of tenure. After setting the other explanatory variables in Equation 1 to their mean, the predicted participation rate increases from 26% ($\sigma = 0.03$) for individuals experiencing the lowest level of police contact to 0.38% ($\sigma = 0.10$) experiencing the highest. These results are plotted in Figure 2.8.[22]

[22] We compare the marginal effect of arrests for marijuana to those for murder in Appendix Figures 2.F.1 and 2.F.2.

Table 2.1: Models of Protest Participation Using Social Activity Hub

	Model I		Model II	
	Coef.	S.E.	Coef.	S.E.
Exposure to Policing†	0.10*	0.04	0.12*	0.05
Proportion Black	0.12	0.31	0.12	0.31
Proportion Hispanic	−2.59	1.79	−2.58	1.79
Proportion Female	−0.60	1.62	−0.60	1.62
Youth Bulge	0.73	0.83	0.73	0.83
Median Income	−1.06	0.92	−1.06	0.92
Poverty Rate	0.00	0.00	0.00	0.00
Unemployment Rate	0.14	0.74	0.15	0.74
Proportion HS Graduate	0.02	1.23	0.00	1.23
Per 1000 Population	0.22	0.25	0.22	0.25
Per 1000 Population^2	−0.05	0.04	−0.05	0.04
Intercept	−0.10	1.51	−0.09	1.51
N	1,239		1,239	
AIC	1478.3		1478.3	

* $p < 0.05$
† Exposure to policing is per 100 arrests in Model I and per 100 visits in Model II.

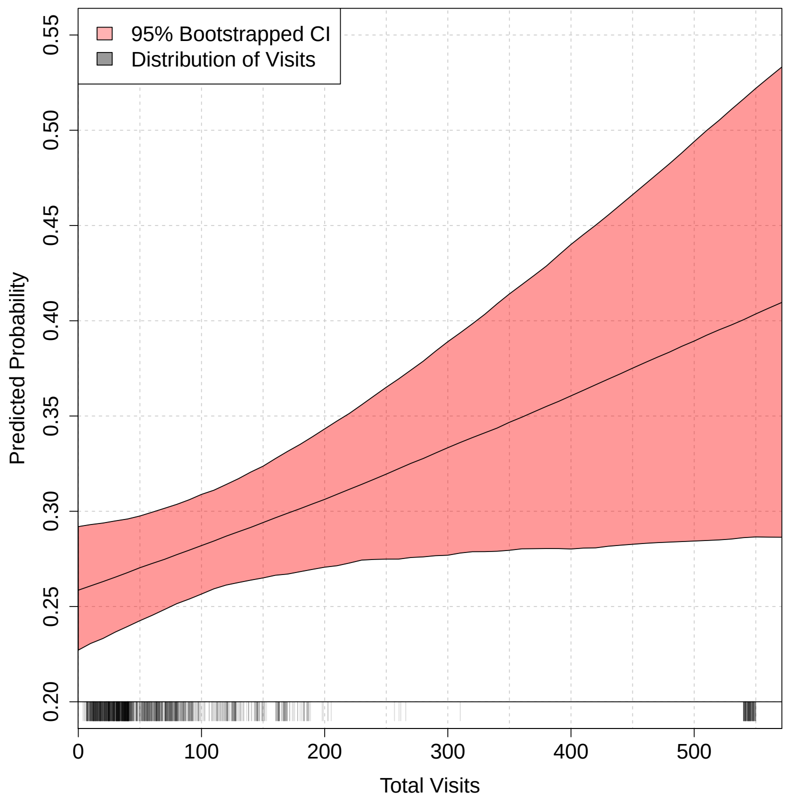

Figure 2.8: Predicted probability of being classified as a protest account simulated using estimated parameters from Model II (i.e. exposure to police is measured by total visits and home locations are estimated using the DPM). 95% confidence interval is obtained using the bootstrap method.

2.4.1 Mechanism

Our theory suggests that encounters with street-level bureaucrats produce grievances, which make participation in collective action more likely. Not all arrests produce grievances in the community. Lipsky (2010) argues that civil servants produce grievances when they exhibit discretion and preferential treatment for some group. In terms of arrests, this observation suggests that the only arrests that generate grievances are those where officers could have shown discretion and leniency but chose not to. This is similar to the negative effect on perceptions of government generated by traffic stops and stop-and-frisk, which rely on officers' subjective judgments (Kang and Dawes 2017; Lerman and Weaver 2014). We do not expect arrests for crimes where officers cannot show discretion to produce grievances.

In our data, grievance-generating arrests include those for marijuana and other drugs, because only a small portion of either drug users and sellers are ever arrested. In contrast, we do not expect arrests for murder or gun crime to generate grievances in the community because most agree severe crime deserves to be punished.

In Table 2.2, we subset our data on arrests according to what the BPD recorded as the initial charge. While the specifics of the charging document might change in between the time the person is first arraigned and is brought to trial, it is a useful proxy of the type of crime in which the arrestee was engaged. We use the 2017 Maryland Sentencing Guidelines Offense Table to code each crime according to its maximum term; offense type; and whether it involves marijuana or guns. Table 2.2 shows our estimated effect of drug arrests on participation in the Freddie Gray protests.

As predicted, we find a strong positive association between discretionary arrests such as arrests for drugs and marijuana in a community with participation in collective action. In contrast, Table 2.3 the opposite. Examining arrests for murder and arrests with a maximum charge of life in prison, we find no association with protest participation.

Table 2.2: Models of Protest Participation Using Social Activity Hub: Drug Arrests

	All Drugs				Marijuana			
	Model I		Model II		Model I		Model II	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
Exposure to Policing†	0.19	0.08	0.21	0.09	2.85	0.90	3.39	1.07
Proportion Black	0.03	0.30	0.04	0.30	-0.01	0.30	-0.01	0.30
Proportion Hispanic	-2.99	1.73	-2.98	1.73	-2.70	1.75	-2.73	1.75
Proportion Female	-0.36	1.56	-0.39	1.56	-0.25	1.55	-0.27	1.55
Youth Bulge	0.43	0.82	0.44	0.82	0.20	0.83	0.22	0.83
Unemployment Rate	-1.35	0.91	-1.35	0.91	-1.59	0.92	-1.58	0.92
Median Income	-0.00	0.00	-0.00	0.00	-0.00	0.00	-0.00	0.00
Poverty Rate	-0.05	0.73	-0.03	0.73	0.01	0.73	0.02	0.73
Proportion HS Graduate	-0.02	1.19	-0.08	1.19	-0.11	1.19	-0.17	1.19
Per 1000 Population	0.23	0.25	0.24	0.25	0.16	0.25	0.16	0.25
Per 1000 Population^2	-0.05	0.03	-0.05	0.03	-0.04	0.03	-0.04	0.03
Intercept	0.09	1.48	0.15	1.48	0.39	1.49	0.42	1.49

† Exposure to policing is per 100 arrests in Model I and per 100 visits in Model II.

Table 2.3: Models of Protest Participation Using Social Activity Hub: Non-Discretionary Arrests

	Murder				Life Sentence			
	Model I		Model II		Model I		Model II	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
Exposure to Policing†	1.14	0.57	1.34	0.70	0.44	0.42	0.48	0.48
Proportion Black	-0.33	0.27	-0.33	0.27	0.04	0.30	0.04	0.30
Proportion Hispanic	-2.76	1.42	-2.76	1.42	-3.10	1.75	-3.11	1.75
Proportion Female	0.41	1.51	0.39	1.51	-0.75	1.59	-0.76	1.60
Youth Bulge	-0.27	0.75	-0.25	0.75	0.81	0.80	0.82	0.80
Unemployment Rate	0.98	0.82	1.00	0.82	-1.14	0.90	-1.13	0.90
Median Income	-0.00	0.00	-0.00	0.00	-0.00	0.00	-0.00	0.00
Poverty Rate	-0.26	0.71	-0.25	0.71	0.13	0.73	0.13	0.73
Proportion HS Graduate	-1.48	1.13	-1.47	1.13	-0.14	1.24	-0.13	1.25
Per 1000 Population	0.07	0.21	0.07	0.21	0.35	0.25	0.36	0.25
Per 1000 Population^2	-0.02	0.03	-0.02	0.03	-0.07	0.03	-0.07	0.03
Intercept	0.35	1.38	0.33	1.38	0.11	1.48	0.10	1.48

† Exposure to policing is per 100 arrests in Model I and per 100 visits in Model II.

2.5 Instrumental Variable

We use a raster of the elevation in Baltimore with data from the State of Maryland's Geographic Information Office to construct an instrumental variable for police encounters. Given the history of Baltimore's urbanization (see Section 2.5.1), black residents were forced to live at lower elevations due to increased disease risk from cholera. Wealthy white residents moved into the hills surrounding the city to escape disease. This results in a marked pattern wherein the Baltimore Police Department makes more arrests today at lower elevations.

2.5.1 IV Assumptions

First Stage Relationship

An instrumental variable (IV) estimation strategy requires that elevation correlate with police behavior prior to the Freddie Grey protests. This relationship is well established, based on the history of settlement in Baltimore and the late introduction of sewers and urban sanitization.

Upon Baltimore's founding in 1729, the initial settlement clustered around the tidal basin of the Patapsco River. Baltimore remained a relatively small settlement compared to nearby Philadelphia until the 1830s, when it became a granary for sugar producing Caribbean colonies. The city also profited from a thriving slave trading market located near Downtown. As Pietila (2010: 8) notes, "in its early history, Baltimore had expanded without much racial fuss. Main streets and alleys were racially, ethnically, and economically mixed."

The formerly mixed city began to experience residential segregation by 1835, when whites had taken over the housing along the city's main thoroughfares (Pietila 2010: 8). Black residents were forced to live in shacks located along alleys without sanitation. Between 1880 and 1900, African-Americans freed from slavery and escaping Reconstruction moved to cities in the north in large numbers. This increased Baltimore's black population by 47% in two decades (Power 1983: 290). This migration, coupled with the Depression of 1882-85, left new

African-American Baltimoreans with few job opportunities. As a result, they "sought out the cheapest housing. . . [and] rented shanties and doubled up in small houses" (Power 1983: 290).

Black Baltimoreans lived in squalid conditions in alleys behind wealthy whites. Described in contemporaneous *Baltimore News* article areas of Baltimore had:

> Open drains, great lots filled with high weeds, ashes and garbage accumulated in the alleyways, cellars filled with filthy black water, houses that are total strangers to the touch of whitewash or scrubbing brush, human bodies that have been strangers for months to soap and water, villainous looking negroes who loiter and sleep around the street corners and never work; vile and vicious women, with but a smock to cover their black nakedness, lounging in the doorways or squatting upon the steps, hurling foul epithets at every passerby; foul streets, foul people, in foul tenements filled with foul air (Crooks 1968: 20).

By this period, Baltimore had become America's seventh largest city and its largest without a sewer system (Pietila 2010; Crooks 1968). The unsanitary conditions resulted in numerous typhus, yellow fever, and cholera outbreaks (Pietila 2010: 9). The development of horse-drawn trolleys and electric streetcars in particular in 1890 had a profound impact on Baltimore (Pietila 2010: 9). Neighbors who could afford to moved away from the low-lying areas around the harbor. Benefitting from the large, hilly estates that historically surrounded the city, Baltimore expanded outwards rapidly and suburbanized. Baltimore is home to some of the first planned communities in America, such as Roland Park (Glotzer 2015).

Racial covenants closed these new "streetcar suburbs" in the hills surrounding the low-lying harbor area to non-white and Jewish residents. As Power (2002: 58) notes, "to protect the white community from invasion, crime, and contagion, [Baltimore] put in place [policies] to isolate" African-American neighborhoods. This culminated in 1910 with the Baltimore City Council's passage of one of the most strict racial residential segregation laws in the country. The law prevented African-Americans from "mov[ing] into a block in which more than half the residents are white," while also preventing whites from "mov[ing] into a block in which more than half the residents are colored" (Power 1983: 299). This law, and similar laws that followed, cemented a pattern wherein wealthy and white residents live in the hills white poor and minority residents live at low elevations on the tidal basin.

2.5.2 IV Results

Table 2.4 shows the estimation results. In all cases, the instrumental variable (an indicator for census tracts above 80 feet) is negatively associated with both policing contact variables.[23] The instrument, moreover, is strong (Staiger and Stock 1997).[24]

The second stage of the estimation is shown in the bottom panel. It shows the local average treatment effect (LATE) using our instrumented measure of policing contact. The bottom panel shows that there is a strong positive association between the policing contact and participation in collective action.

2.6 Robustness

The results in the previous section show that there is a strong, positive relationship between police activity and participation in collective action. In our online appendix, we report a variety of alternate specifications to assess this relationship's robustness. In Appendix Table 2.D.1, we confirm our results do not depend upon using social activity hubs to measure political context. Instead of using this novel measure of context, we use each user's modal tweet location. This alternate measure does not substantively change our results.

Table 2.E.1 explores whether the evidence for our mechanism is robust to alternate types of crimes. In the appendix, we show that arrests for crimes defined by the State of Maryland as "personal" (i.e. kidnapping and assault) or crimes where the perpetrator was armed have no effect on participation in collective action. This further confirms our mechanism that only discretionary arrests increase participation.

[23] We select 80 feet because it is the median value of our elevation measure. Our results do not depend upon this choice of elevation.

[24] F-tests on the instrument, shown in the table, all exceed the rule of thumb value of 10 for strong instruments.

Table 2.4: IV Regression Estimation Results

	First Stage			
	DV: Exposure to Policing			
	Model I		Model II	
	Coef.	S.E.	Coef.	S.E.
Elevation$_{(80')}$	−0.94	(0.12)	−0.80	(0.10)
N	1,303		1,303	
R^2	0.05		0.05	
F-Statistic on Instrument	204.65		197.67	
	Second Stage			
	DV: Protest Participation			
	Model I		Model II	
	Coef.	S.E.	Coef.	S.E.
Exposure to Policing†	0.09	(0.03)	0.10	(0.04)
N	1,303		1,303	

† Exposure to policing is per 100 arrests in Model I and per 100 visits in Model II.
Regressions include a constant, not shown.

Model dependence is an additional concern we address in our online appendix. We estimate a least absolute shrinkage and selection operator (Lasso) regression to select only the subset of variables that minimize prediction error. These results, presented in Table 2.G.1, show that the Lasso selects both measures of contact with police. We validate these variable selection results with a random forest. These results, presented in Table 2.H.3, show our measure of policing contact have the highest variable importance scores.

In another robustness check, we consider whether outliers in our data bias our results. We estimate the Cook's distance and DFBeas for the data on our model, and find little evidence of outliers. This exercise, summarized in Figures 2.H.1 and 2.H.2, suggests our results are not the product of influential observations or outliers.

2.7 Discussion

The theory presented in this paper predicts that variation in grievance intensity can help explain participation in collective action. We argue that political context is an important — and overlooked — source of grievances (Enos 2014). Prior theoretical accounts of the link between grievance and costly political participation focuses on horizontal inequalities (Gurr 1970; Cederman, Weidmann, and Gleditsch 2011). Even if inequality between groups is informative regarding *which* groups mobilize, it does not explain why some people participate while others abstain. Building on Lipsky (2010), we contend that interactions with "street-level bureaucrats" is both contextually-dependent and an important source of grievances.

Data limitations have hitherto limited our ability to test conjectures about individual-level variation in grievances or protest participation; measuring whether an individual participated in a protest or their relevant grievances is non-trivial. Prior studies address this issue by either surveying protestors in the field (Lawrence 2016; Aytaç, Schiumerini, and Stokes 2017) or using observational designs (Madestam et al. 2013; Kern 2011). While field surveys yield measures participation, desirability bias might influence respondents' answers. Individual-level inferences from aggregate data in observational studies, moreover, might be biased.

We argue that it is essential to measure context at the individual-level in order to understand participation in collective action. To do so, we exploit social media data from the protests during April 2015 in Baltimore, MD after the death of Freddie Gray in police custody. After Gray was arrested, he fell into a coma in a police van and later died from his injuries. This triggered a wave of protests against police brutality throughout the city that only ended after the Maryland National Guard imposed a curfew. While participation was widespread, it was not complete. Drawing on the literature on the effect of encountering the carceral state, we argue that interacting with the police generate more intense grievances (White 2015; Uggen and Manza 2002). In turn, individuals subject to more police activity were more likely to protest.

We surmount earlier data limitations by purchasing every geotagged tweet made within Baltimore in April and May 2015. We then train a classifier model to identify whether every account belongs to a protestor. Despite sparse data, our classifier model has an accuracy rate of approximately 95%. We then purchase the users' entire Twitter history and pass the geotagged tweets through a novel community detection algorithm. Within the data, each community is a different area in Baltimore where the user spends time (Rossmo 1999; Verity et al. 2014). These communities, which we refer to as social activity hubs (SAHs), are a novel estimation strategy to measure context from social media data.

We find a strong positive relationship between police activity and participation in collective action. This finding has a number of implications for our understanding of political behavior and highlights opportunities for additional research. First, prior research on the carceral state strongly suggest interactions with police officers demobilize voters (Bruch, Ferree, and Soss 2010; Hjalmarsson and Lopez 2010; Burch 2011; Meredith and Morse 2015; Gerber et al. 2015; White 2015). Our results suggest police encounters have a displacing effect on political participation. Rather than decrease total levels of participation, police activity might instead displace participation into informal channels such as activism or protest. Linking social media accounts to voter files would be useful to test this conjecture.

Chapter 2 in part is currently being prepared for submission for publication of the material. Chen, Ted Hsun Yun; Christopher J. Fariss; and Paul Zachary.

Bibliography

Acemoglu, Daron, and James A Robinson. 2006. "Economic backwardness in political perspective." *American Political Science Review* 100 (1): 115–131.

Aytaç, S Erdem, Luis Schiumerini, and Susan Stokes. 2017. "Why Do People Join Backlash Protests? Lessons from Turkey." *Journal of Conflict Resolution* .

Blinder, Alan, and Richard Pérez-Pe na. 2015. "6 Baltimore Police Officers Charged in Freddie Gray Death." *New York Times* .

Boix, Carles. 2008. "Economic roots of civil wars and revolutions in the contemporary world." *World Politics* 60 (3): 390–437.

Bruch, Sarah K, Myra Marx Ferree, and Joe Soss. 2010. "From policy to polity: Democracy, paternalism, and the incorporation of disadvantaged citizens." *American Sociological Review* 75 (2): 205–226.

Buhaug, Halvard, Lars-Erik Cederman, and Jan Ketil Rød. 2008. "Disaggregating ethnonationalist civil wars: A dyadic test of exclusion theory." *International Organization* 62 (3): 531–551.

Burch, Traci. 2011. "Turnout and party registration among criminal offenders in the 2008 general election." *Law & Society Review* 45 (3): 699–730.

Burch, Traci. 2013. *Trading democracy for justice: Criminal convictions and the decline of neighborhood political participation*. Chicago, IL: University of Chicago Press.

Butler, Daniel M, and David E Broockman. 2011. "Do politicians racially discriminate against constituents? A field experiment on state legislators." *American Journal of Political Science* 55 (3): 463–477.

Carrington, Keith. 2005. "Is there a need for control?" *Public Administration Quarterly* pp. 140–161.

Carrubba, Clifford J. 2005. "Courts and compliance in international regulatory regimes." *Journal of Politics* 67 (3): 669–689.

Cederman, Lars-Erik, Andreas Wimmer, and Brian Min. 2010. "Why do ethnic groups rebel? New data and analysis." *World Politics* 62 (1): 87–119.

Cederman, Lars-Erik, Nils B Weidmann, and Kristian Skrede Gleditsch. 2011. "Horizontal inequalities and ethnonationalist civil war: A global comparison." *American Political Science Review* 105 (3): 478–495.

Chen, Ted, Chris Fariss, and Paul Zachary. 2017. "Social Activity Hubs (SAHs): A New Method for Estimating User Specific Contextual Factors From Social Media Data. In *Computational Social Science 2017*, ed. David Krakauer, and Scott Page.

Chong, Dennis. 2014. *Collective action and the civil rights movement*. Chicago, IL: University of Chicago Press.

Clack, B., J. Dixon, and C. Tredoux. 2005. "Eating together apart: Patterns of segregation in a multi-ethnic cafeteria." *Journal of Community and Applied Social Psychology* 15: 1–16.

Collier, Paul, and Anke Hoeffler. 2004. "Greed and grievance in civil war." *Oxford economic papers* 56 (4): 563–595.

Crooks, James B. 1968. *Politics & progress: The rise of urban progressivism in Baltimore, 1895 to 1911*. Baton Rouge, LA: Louisiana State Univeristy Press.

Davenport, Christian. 2005. "Understanding Covert Repressive Action: The Case of the U.S. Government against the Republic of New Africa." *Journal of Conflict Resolution* 49 (1): 120 – 140.

Davenport, Christian. 2007. "State repression and political order." *Annual Review of Political Science* 10: 1–23.

Davenport, Christian. 2010. *Media Bias, Perspective, and State Repression: The Black Panther Party*. Cambridge, MA: Cambridge University Press.

Davenport, Christian, Sarah A. Soule, and David A. Armstrong. 2011. "Protesting While Black? The Differential Policing of American Activism, 1960 to 1990." *American Sociological Review* 76 (1): 152–178.

Earl, Jennifer, Andrew Martin, John D. McCarthy, and Sarah A. Soule. 2004. "The use of newspaper data in the study of collective action." *Annual Review of Sociology* 30: 65–80.

Enos, Ryan D. 2014. "Causal effect of intergroup contact on exclusionary attitudes." *Proceedings of the National Academy of Sciences* 111 (10): 3699–3704.

Enos, Ryan D. 2016. "What the demolition of public housing teaches us about the impact of racial threat on political behavior." *American Journal of Political Science* 60 (1): 123–142.

Enos, Ryan D, and Noam Gidron. 2016. "Intergroup behavioral strategies as contextually determined: Experimental evidence from Israel." *The Journal of Politics* 78 (3): 851–867.

Fearon, James D, and David D Laitin. 2003. "Ethnicity, insurgency, and civil war." *American political science review* 97 (1): 75–90.

Fearon, James D, Kimuli Kasara, and David D Laitin. 2007. "Ethnic minority rule and civil war onset." *American Political science review* 101 (1): 187–193.

Ganz, Marshall. 2009. *Why David sometimes wins: Leadership, organization, and strategy in the California farm worker movement*. New York, NY: Oxford University Press.

Gause, LaGina. 2017. "The Advantage of Disadvantage: Legislative Responsiveness to Collective Action by the Politically Marginalized." Unpublished book manuscript.

Gentry, Jeff. 2015. *twitteR: R Based Twitter Client*. R package version 1.1.9.
URL: *http://CRAN.R-project.org/package=twitteR*

Gerber, Alan S, Gregory A Huber, Marc Meredith, Daniel R Biggers, and David J Hendry. 2015. "Can incarcerated felons be (Re) integrated into the political system? Results from a field experiment." *American Journal of Political Science* 59 (4): 912–926.

Gerring, John. 2008. "Case Selection for Case-Study Analysis: Qualitative and Quantitative Techniques." In *The Oxford handbook of political methodology*, ed. Janet M. Box-Steffensmeier, Henry E. Brady, and David Collier. New York, NY: Oxford University Press.

Glotzer, Paige. 2015. "Exclusion in Arcadia: how suburban developers circulated ideas about discrimination, 1890–1950." *Journal of Urban History* 41 (3): 479–494.

Gurr, Ted Robert. 1970. *Why Men Rebel*. Princeton, NJ: Princeton University Press.

Gurr, Ted Robert. 1993. "Why minorities rebel: A global analysis of communal mobilization and conflict since 1945." *International Political Science Review* 14 (2): 161–201.

Han, Hahrie. 2014. *How organizations develop activists: Civic associations and leadership in the 21st century*. New York, NY: Oxford University Press.

Hauge, Michelle V, Mark D Stevenson, D Kim Rossmo, and Steven C Le Comber. 2016. "Tagging Banksy: Using geographic profiling to investigate a modern art mystery." *Journal of Spatial Science* 61 (1): 185–190.

Hegre, Håvard, and Nicholas Sambanis. 2006. "Sensitivity analysis of empirical results on civil war onset." *Journal of conflict resolution* 50 (4): 508–535.

Hjalmarsson, Randi, and Mark Lopez. 2010. "The voting behavior of young disenfranchised felons: Would they vote if they could?" *American Law and Economics Review* pp. 356–393.

Horowitz, Donald L. 1985. *Ethnic groups in conflict*. Berkeley, CA: Univ of California Press.

Kang, Woo Chang, and Christopher Dawes. 2017. "The Electoral Effect of Stop-and-Frisk." SSRN Working Paper.

Kern, Holger Lutz. 2011. "Foreign media and protest diffusion in authoritarian regimes: The case of the 1989 East German revolution." *Comparative Political Studies* 44 (9): 1179–1205.

King, Gary. 1997. *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*. Princeton, NJ: Princeton University Press.

Kuran, Timur. 1989. "Sparks and prairie fires: A theory of unanticipated political revolution." *Public choice* 61 (1): 41–74.

Lawrence, Adria K. 2016. "Repression and Activism among the Arab Spring's First Movers: Evidence from Morocco's February 20th Movement." *British Journal of Political Science* .

Lerman, Amy E, and Vesla M Weaver. 2014. *Arresting citizenship: The democratic consequences of American crime control*. Chicago, IL: University of Chicago Press.

Lichbach, Mark. 1998. *The Rebel's Dilemma*. Ann Arbor, MI: University of Michigan Press.

Lichbach, Mark I. 1994. "What Makes Rational Peasants Revolutionary? Dilemma, Paradox, and Irony in Peasant Collective Action." *World Politics* 46 (3): 383–418.

Lichbach, Mark Irving. 1989. "An evaluation of "does economic inequality breed political conflict?" studies." *World politics* 41 (4): 431–470.

Lipsky, Michael. 2010. *Street-level bureaucracy: dilemmas of the individual in public service*. New York, NY: Russell Sage Foundation.

Lohmann, Susanne. 1994. "The dynamics of informational cascades: The Monday demonstrations in Leipzig, East Germany, 1989–91." *World politics* 47 (1): 42–101.

Lohmann, Susanne. 1995. "Information, access, and contributions: A signaling model of lobbying." *Public Choice* 85 (3): 267–284.

Lorentzen, Peter. 2014. "China's Strategic Censorship." *American Journal of Political Science* 58 (2): 402–414.

Madestam, Andreas, Daniel Shoag, Stan Veuger, and David Yanagizawa-Drott. 2013. "Do political protests matter? evidence from the tea party movement." *The Quarterly Journal of Economics* p. qjt021.

Mason, T. David. 1984. "Individual Participation in Collective Racial Violence: A Rational Choice Synthesis." *American Political Science Review* 78 (4): 1040–1056.

Mason, T. David, and Dale A. Krane. 1989. "The Political Economy of Death Squads: Toward a Theory of the Impact of State-Sanctioned Terror." *International Studies Quarterly* 33 (2): 175–198.

Matthews, Brian W. 1975. "Comparison of the predicted and observed secondary structure of T4 phage lysozyme." *Biochimica et Biophysica Acta (BBA)-Protein Structure* 405 (2): 442–451.

Meredith, Marc, and Michael Morse. 2015. "The politics of the restoration of ex-felon voting rights: The case of Iowa." *Quarterly Journal of Political Science* 10.

Miguel, Edward, Shanker Satyanath, and Ernest Sergenti. 2004. "Economic shocks and civil conflict: An instrumental variables approach." *Journal of political Economy* 112 (4): 725–753.

Murdie, Amanda, and Tavishi Bhasin. 2011. "Aiding and abetting: Human rights INGOs and domestic protest." *Journal of Conflict Resolution* 55 (2): 163–191.

Naveed, Nasir, Thomas Gottron, Jérôme Kunegis, and Arifah Che Alhadi. 2011. Bad news travel fast: A content-based analysis of interestingness on twitter. In *Proceedings of the 3rd International Web Science Conference*. ACM p. 8.

Neal, Radford M. 2000. "Markov chain sampling methods for Dirichlet process mixture models." *Journal of computational and graphical statistics* 9 (2): 249–265.

Norris, Pippa, Stefaan Walgrave, and Peter Van Aelst. 2005. "Who demonstrates? Antistate rebels, conventional participants, or everyone?" *Comparative politics* pp. 189–205.

O'Leary, Mike. 2010. Implementing a Bayesian approach to criminal geographic profiling. In *COM. Geo*.

Olson, Mancur. 1965. *The Logic of Collective Action*. Cambridge, MA: Cambridge University Press.

Ostrom, Elinor. 1990. *Governing the Commons*. Cambridge, MA: Cambridge Cambridge University Press.

Pavalanathan, Umashanthi, and Jacob Eisenstein. 2015. Confounds and Consequences in Geotagged Twitter Data. In *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*.

Payne, Charles M. 2007. *I've got the light of freedom: The organizing tradition and the Mississippi freedom struggle*. Berkeley, CA: University of California Press.

Pietila, Antero. 2010. *Not in My Neighborhood: How Bigotry Shaped a Great American City*. Lanham, MD: Rowman & Littlefield.

Platt, John. 1999. "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods." *Advances in large margin classifiers* 10 (3): 61–74.

Popkin, Samuel L. 1979. *The Rational Peasant: The Political Economy of Rural Society in Vietnam*. Berkeley, CA: University of California Press.

Popkin, Samuel L. 1988. "Political entrepreneurs and peasant movements in Vietnam." *Rationality and revolution* pp. 9–62.

Power, Garrett. 1983. "Apartheid Baltimore style: The residential segregation ordinances of 1910-1913." *Maryland Law Review* 42: 289–328.

Power, Garrett. 2002. *From Mobtown to Charm City*. Baltimore, MD: Maryland Historical Society chapter Deconstructing the Slums of Baltimore.

Regan, Patrick M, and Daniel Norton. 2005. "Greed, grievance, and mobilization in civil wars." *Journal of Conflict Resolution* 49 (3): 319–336.

Robinson, James A, and Ragnar Torvik. 2005. "White elephants." *Journal of Public Economics* 89 (2): 197–210.

Ross, Michael L. 2004. "What do we know about natural resources and civil war?" *Journal of peace research* 41 (3): 337–356.

Rossmo, D Kim. 1999. *Geographic profiling*. CRC press.

Saif, Hassan, Yulan He, Miriam Fernandez, and Harith Alani. 2014. Semantic patterns for sentiment analysis of Twitter. In *International Semantic Web Conference*. Springer pp. 324–340.

Sambanis, Nicholas. 2001. "Do ethnic and nonethnic civil wars have the same causes? A theoretical and empirical inquiry (part 1)." *Journal of Conflict Resolution* 45 (3): 259–282.

Sands, Melissa L. 2017. "Exposure to inequality affects support for redistribution." *Proceedings of the National Academy of Sciences* p. 201615010.

Schneider, Mark, and Paul Teske. 1992. "Toward a Theory of the Political Entrepreneur: Evidence From Local Government." *American Political Science Review* 86 (03): 737–747.

Schussman, Alan, and Sarah A Soule. 2005. "Process and protest: Accounting for individual protest participation." *Social forces* 84 (2): 1083–1108.

Smith, Douglas A. 1986. "The neighborhood context of police behavior." *Crime and justice* 8: 313–341.

Snyder, David, and Charles Tilly. 1972. "Hardship and collective violence in France, 1830 to 1960." *American Sociological Review* pp. 520–532.

Staiger, Douglas, and James H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65 (3): 557–586.
URL: *http://www.jstor.org/stable/2171753*

Steinert-Threlkeld, Zachary C. 2017. "Spontaneous Collective Action: Peripheral Mobilization During the Arab Spring." *American Political Science Review* .

Steinert-Threlkeld, Zachary C., Delia Mocanu, Alessandro Vespignani, and James H. Fowler. 2015. "Online Social Networks and Offline Protest." *EPJ Data Science* 4 (19): 1–9.

Stevenson, M.D, Verity, and R. 2014. *Rgeoprofile : Geographic Profiling in R*. London, England: Queen Mary University of London. Version 1.2.
URL: *http://evolve.sbcs.qmul.ac.uk/lecomber/sample-page/geographic-profiling/*

Stewart, Frances. 2008. *Horizontal Inequalities and Conflict: An Introduction and Some Hypotheses*. Palgrave Macmillan chapter Horizontal Inequalities and Conflict: An Introduction and Some Hypotheses, pp. 3–24.

Tarrow, Sidney. 1994. *Power in movement: Social movements, collective action and mass politics*. New York, NY: Cambridge University Press.

Thomson, Henry. 2016. "Rural Grievances, Landholding Inequality, and Civil Conflict." *International Studies Quarterly* 60 (3): 511–519.

Tullock, Gordon. 1971. "The Paradox of Revolution." *Public Choice* 11: 89–99.

Tyler, Tom R, Jeffrey Fagan, and Amanda Geller. 2014. "Street stops and police legitimacy: Teachable moments in young urban men's legal socialization." *Journal of Empirical Legal Studies* 11 (4): 751–785.

Uggen, Christopher, and Jeff Manza. 2002. "Democratic contraction? Political consequences of felon disenfranchisement in the United States." *American Sociological Review* pp. 777–803.

Verity, Robert, Mark D Stevenson, D Kim Rossmo, Richard A Nichols, and Steven C Le Comber. 2014. "Spatial targeting of infectious disease control: identifying multiple, unknown sources." *Methods in Ecology and Evolution* 5 (7): 647–655.

Walls, David S. 2015. *Community organizing*. San Francisco, CA: John Wiley & Sons.

White, Ariel. 2015. "Misdemeanor Disenfranchisement? The demobilizing effects of brief jail spells on potential voters." Harvard Scholar.

White, Ariel R, Noah L Nathan, and Julie K Faller. 2015. "What do I need to vote? Bureaucratic discretion and discrimination by local election officials." *American Political Science Review* 109 (1): 129–142.

Wildeman, Christopher. 2014. "Parental incarceration, child homelessness, and the invisible consequences of mass imprisonment." *The Annals of the American Academy of Political and Social Science* 651 (1): 74–96.

Willer, Robb. 2009. "Groups reward individual sacrifice: The status solution to the collective action problem." *American Sociological Review* 74 (1): 23–43.

Williamson, Vanessa, Theda Skocpol, and John Coggin. 2011. "The Tea Party and the remaking of Republican conservatism." *Perspectives on Politics* 9 (1): 25–43.

Wood, Elisabeth Jean. 2003. *Insurgent collective action and civil war in El Salvador*. New York, NY: Cambridge University Press.

Zuern, Elke. 2011. *The Politics of Necessity: community organizing and democracy in South Africa*. Madison, WI: University of Wisconsin Press.

Appendix

2.A Classification

2.A.1 Preprocessing

Because Twitter imposes a limit of 140 characters on all tweets, each tweet's brevity poses unique inference problems (Saif et al. 2014; Naveed et al. 2011).Users frequently abbreviate or use slang to maximize the content of each tweet. This implies there can be substantial variation in language use patterns, even among Twitter users discussing the same topic. Such variation, coupled with the brevity of each tweet, results in highly-dimensional, sparse data. As there may not be overlap in the context of tweets among users, this variation poses problems for text analysis algorithms. Insufficient overlap is a problem, because it means our training set will be uninformative, resulting in low accuracy scores.

To address this problem, we preprocess our data extensively prior to analysis. First, we use the CamelCase package in Python to split hashtags into separate words whenever possible.[25] Second, we increase the textual overlap in our corpus by stemming and normalizing our corpus with NLTK's TweetTokenizer package. By stemming our corpus, we reduce inflected words to their word stem.[26] In order to normalize our corpus and reduce dimensionality, we convert emoji into unicode; preserve punctuation; and remove accents from letters.

[25]For instance, the common hashtag #FreddieGrayProtest became "Freddie Gray protest" after processing.

[26]A word stem is the grammatically uninflected root word. For example, "paste" is the stem of words such as "pastes," "pasted," and "pasting."

Finally, after normalizing our corpus, we tokenize our hand coded training set and obtain feature vectors. We obtain features using term-frequency inverse-document frequency (TF-IDF). While other tokenization algorithms weigh every feature's importance by counting the number of times it appears in a corpus, TF-IDF improves upon this approach accounting for both the number of times each feature appears and the total number of words in the corpus. We allow our $n-$gram size to range from 1 to 5 and converted to binary with a threshold of 0. We then convert these features into a vector in which each element represents a single feature (obtained via TF-IDF) and its value is its weight. In order words, we transform every tweet into a vector in which each unique word is an element and that element's value is the frequency that word appears in our corpus divided by the weight.

2.B Figures

2.B.1 ROC Plot

As Figure 2.B.1 shows, the false positive rates for our LR and SVC models are similar.

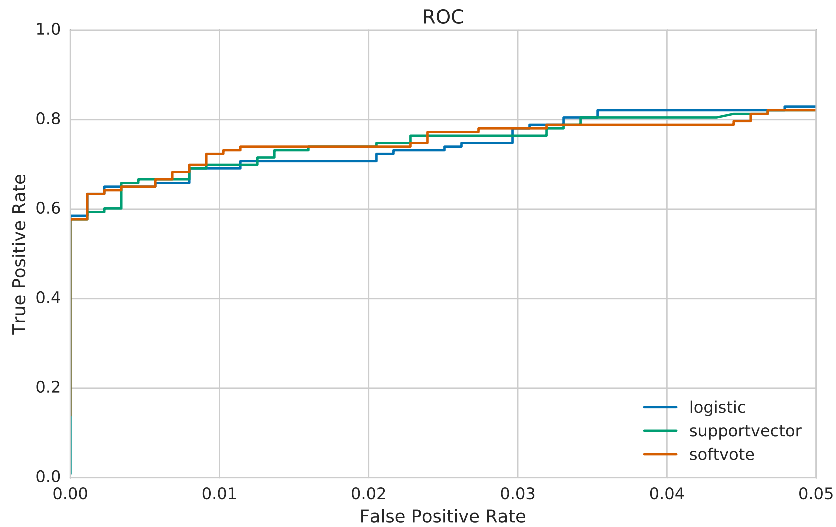

Figure 2.B.1: This plot shows the receiver operating characteristic (ROC) curve from the three estimated models. When using a soft vote model with a false positive rate of 0.01, our true positive rate is approximately 80%.

2.B.2 Features

Figure 2.B.2 graphs the features from our hand-coded training set via t-distributed stochastic neighbor embedding (t-SNE). The t-SNE algorithm maps the high-dimensional feature space into two-dimensional space while trying to preserve the local distance structure for visualization. Although the axes are unintereptable, t-SNE plots are useful to visualize the amount of overlap among features. Appendix Figure 2.B.2 suggests that the features from protest-related and non-protest related twitter accounts are not immediately separable, motivating our use of a classifier algorithm.

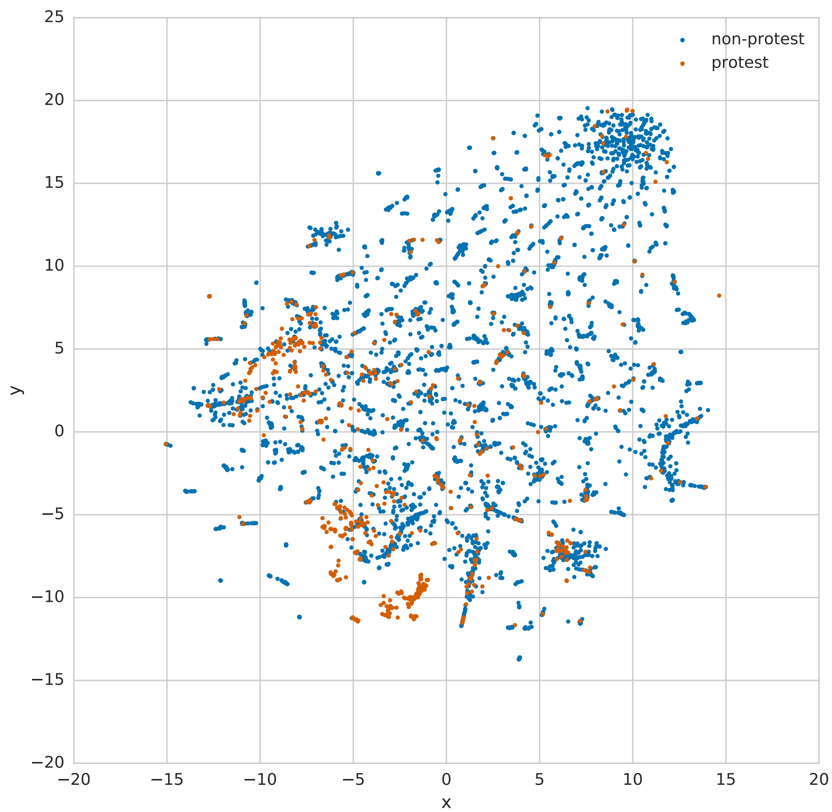

Figure 2.B.2: t-SNE plot that collapses the high-dimensional feature matrix into two dimensions. Features from protest-related accounts are denoted in orange, while non-protest related accounts are denoted in blue.

2.B.3 Density of Control Variables

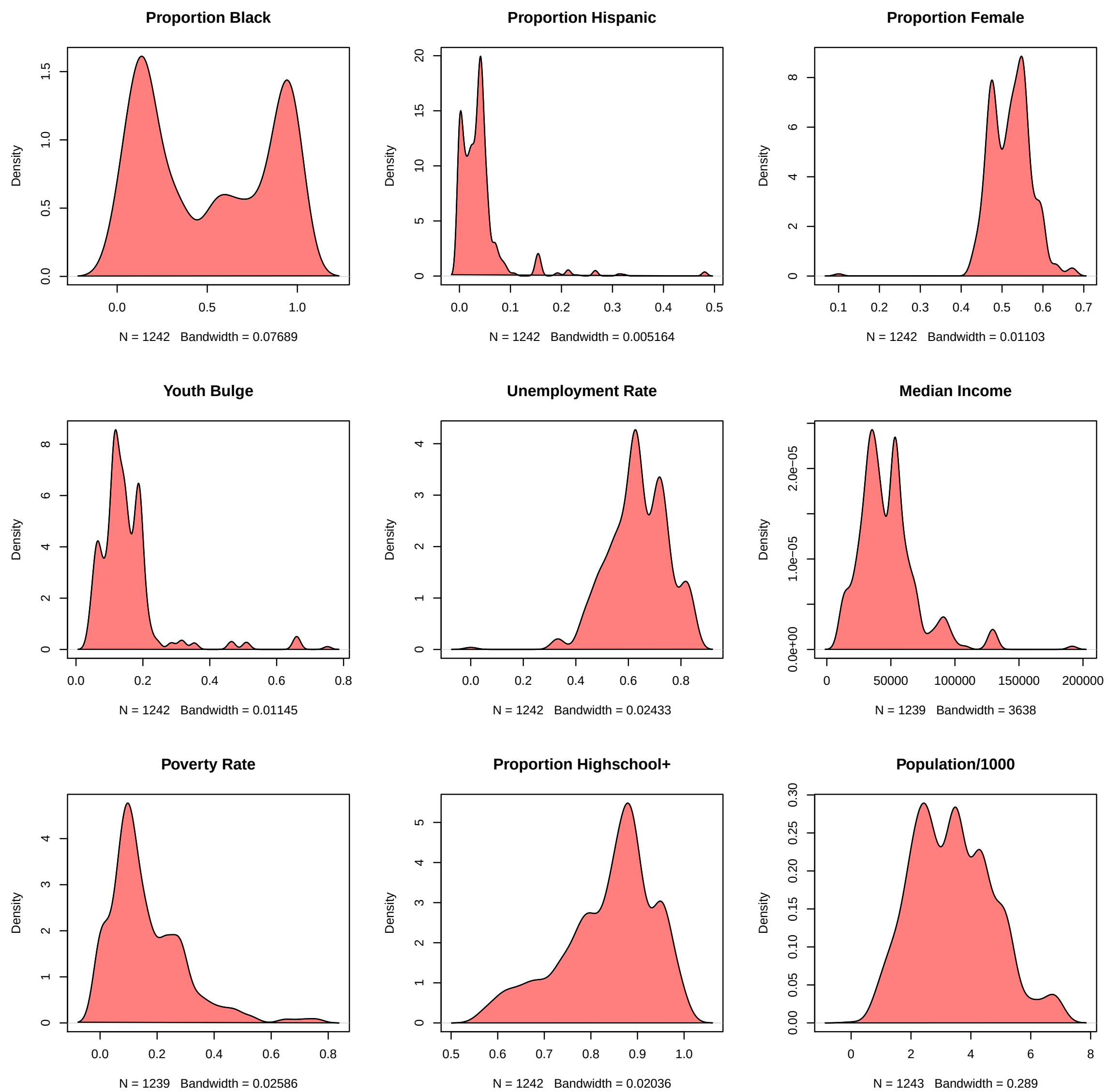

Figure 2.B.3: Density of the control variables in the DPM models.

2.B.4 Scatterplots of Police Activity by Race

These figures show scatterplots of police activity (Y-axis) and percentage African-American (X-axis). While they show a relationship between race and policing, they also show that high proportion African-American census tracts have high and low police contact.

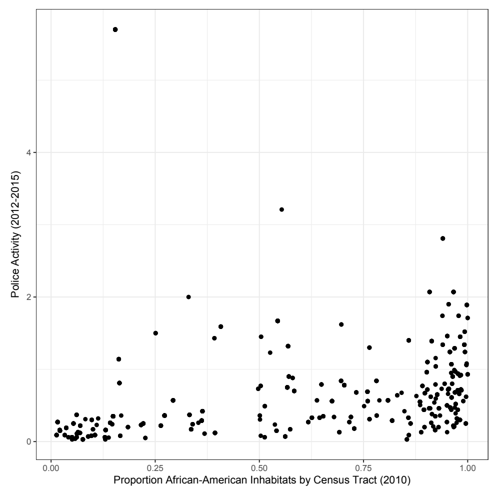

Figure 2.B.4: Scatterplot of arrests by percentage African-American.

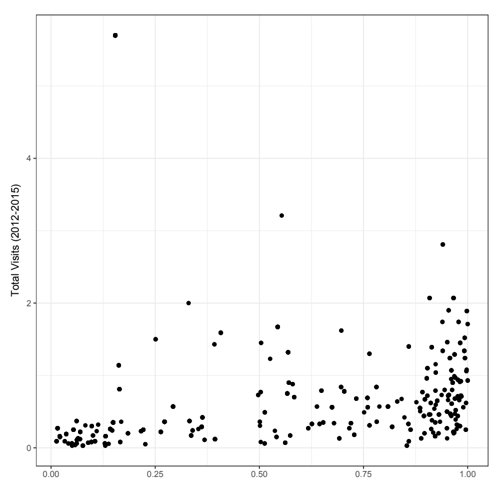

Figure 2.B.5: Scatterplot of total visits by percentage African-American.

2.C Algoritms

2.C.1 Algorithm I

Algorithm 1: Social Activity Hub Estimation for Each User

Data: The set of n observed incidents $\mathbf{x}_i = (x_i^{(1)}, x_i^{(2)})$, $i = 1, ..., n$

if $n = 1$ **then**
    Assign $\mathbf{x}_1$ as $\text{SAH}^{(\text{sole observed})}$;
else
    MCMC algorithm implemented as the `RunMCMC()` function in Rgeoprofile 1.2 (Stevenson et al. 2014: summarized below in algorithm 4), based on the discussion in Verity et al. (2014));
    if *convergence fails;*
    then
        if $n > 25$ **then**
            remove incidents outside of bounding box created by the 1st and 99th percentiles of $x^{(1)}$ and $x^{(2)}$;
        Assign $\mathbf{x}_i$ closest in Euclidean distance to the spatial mean of $\mathbf{x}$ as $\text{SAH}^{(\text{spatial mean})}$;
    else
        Take 2000 posterior draws; thin by keeping the first of every 20;
        begin local minima model:
            combine all 100 posterior draws;
            a) calculate hit score surface;
            b) find local minima j on surface within 0.05 degree decimal radius, $j = 1, ..., \infty$;
            c) **foreach** *local minimum* j **do**
                Assign $\mathbf{x}_i$ closest in Euclidean distance to local minimum as $\text{SAH}_j^{(\text{local minima})}$
            end
            foreach *posterior draw* **do**
                a to c;
            end
        end
        begin cluster mean model:
            combine all 100 posterior draws;
            d) **foreach** *cluster* j *of* $\mathbf{x}$ **do**
                Assign $\mathbf{x}_i$ closest in Euclidean distance to estimated source of cluster as $\text{SAH}_j^{(\text{cluster mean})}$
            end
            foreach *posterior draw* **do**
                d;
            end
        end
    end
end

Result: **SAH**=$(\text{SAH}^{(\text{sole observed})}, \text{SAH}^{(\text{spatial mean})}, \text{SAH}^{(\text{local minima})}, \text{SAH}^{(\text{cluster mean})})$

2.C.2 Algorithm II

Algorithm 2: `RunMCMC()` from Rgeoprofile 1.2 (Stevenson et al. 2014)

Data: The set of n observed incidents $\mathbf{x}_i = (x_i^{(1)}, x_i^{(2)})$, $i = 1, \ldots, n$

Initialize by setting initial values and computing relevant priors;

Define sampling steps:

a) draw and update $\mathbf{z}_{c_i}$ based on most updated c_i;

b) draw and update c_i based on most updated $\mathbf{z}$;

begin Burn-in
  repeat
    for *i in* 1 *to n* **do** a-b;
  until *convergence*;
end

begin Posterior draws
  foreach *posterior draw* **do**
    for *i in* 1 *to n* **do** a-b;
  end
end

Result:

1. For each $\mathbf{x}_{1,\ldots,n}$, its corresponding cluster c_i

2. For each unique cluster c_j, its source location $\mathbf{z}_j$

2.D Alternate Estimation of Social Activity Hubs

In the main body of the paper, we estimate each user's social context using our novel social activity hubs (SAHs) approach. This has the benefit of allowing users' context to range across multiple places within Baltimore. However, it is possible that our results are an artifact of some as-yet unknown bias in this estimation procedure. In Table 2.D.1, we explore this possibility by assigning each user's social activity to their the modal tweet location, i.e. the place from which they tweet most often. Our results in this table are almost entirely consistent with our model estimated using our SAH measure.

Table 2.D.1: Models of Protest Participation; Home Locations by Modal Tweet

	Model III		Model IV	
	Estimate	S.E.	Estimate	S.E.
Exposure to Policing†	0.10*	0.04	0.11*	0.05
Proportion Black	−0.33	0.27	−0.32	0.27
Proportion Hispanic	−2.71	1.41	−2.71	1.41
Proportion Female	0.61	1.51	0.62	1.51
Youth Bulge	−0.30	0.75	−0.31	0.75
Median Income	1.00	0.82	1.01	0.82
Poverty Rate	0.00	0.00	0.00	0.00
Unemployment Rate	−0.33	0.71	−0.32	0.71
Proportion HS Graduate	−1.40	1.11	−1.43	1.11
Per 1000 Population	0.03	0.21	0.03	0.21
Per 1000 Population^2	−0.02	0.03	−0.02	0.03
Intercept	0.16	1.37	0.17	1.37
N	1,501		1,501	
AIC	1863.5		1863.3	

* $p < 0.05$
† Exposure to policing is per 100 arrests in Model III and per 100 visits in Model IV.

2.E Other Non-Grievance Inducing Arrests

In our mechanism section, we argue that arrests where officers can show discretion generate grievances, while arrests supported by the community do not. In the main body, we show that arrests for murder or for crimes that carry a maximum penalty of life do not affect participation in collective action. In this section, we consider whether our evidence behind our mechanism depends upon our choice of the type of crime we consider "community-supported." Specifically, we also explore whether arrests for "personal" crimes such as assault, abuse, and kidnapping affect participation. We also explore whether crimes where a weapon was used affects participation. The results, summarized in Tab 2.E.1, do not suggest our choice of crime affects our results, as arrests are not significant in any model.

Table 2.E.1: Models of Protest Participation Using Social Activity Hub: Drug Arrests

	"Personal" Crimes				Armed			
	Model I		Model II		Model I		Model II	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
Exposure to Policing†	0.14	0.10	0.17	0.11	0.48	0.38	0.57	0.44
Proportion Black	0.07	0.30	0.07	0.30	0.08	0.31	0.07	0.31
Proportion Hispanic	-3.07	1.75	-3.07	1.75	-3.06	1.75	-3.06	1.75
Proportion Female	-0.55	1.59	-0.51	1.59	-0.84	1.55	-0.83	1.55
Youth Bulge	0.74	0.80	0.73	0.80	0.69	0.82	0.67	0.82
Unemployment Rate	-1.15	0.90	-1.14	0.90	-1.24	0.91	-1.25	0.91
Median Income	-0.00	0.00	-0.00	0.00	-0.00	0.00	-0.00	0.00
Poverty Rate	0.09	0.73	0.09	0.73	0.11	0.73	0.10	0.73
Proportion HS Graduate	-0.24	1.23	-0.25	1.23	-0.09	1.21	-0.09	1.21
Per 1000 Population	0.31	0.25	0.31	0.25	0.32	0.25	0.32	0.25
Per 1000 Population^2	-0.06	0.03	-0.06	0.03	-0.06	0.03	-0.06	0.03
Intercept	0.09	1.48	0.07	1.48	0.21	1.49	0.21	1.49

† Exposure to policing is per 100 arrests in Model I and per 100 visits in Model II.

2.F Comparing Marginal Effects for Arrest Types

This section compares the predicted probability of being classified as a protest account using arrests for marijuana and for murder. It finds there is no statistically significant affect for murder arrests, while there is a clear positive association between arrests for marijuana and likelihood of being classified as a protestor.

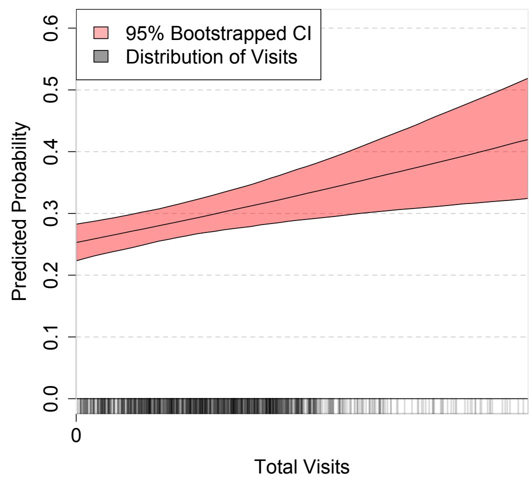

Figure 2.F.1: Predicted probability of being classified as a protest account simulated using estimated parameters from Model II (i.e. exposure to police is measured by total visits for marijuana and home locations are estimated using the DPM). 95% confidence interval is obtained using the bootstrap method.

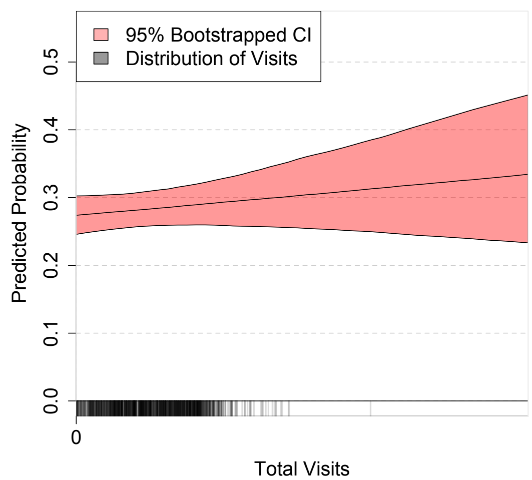

Figure 2.F.2: Predicted probability of being classified as a protest account simulated using estimated parameters from Model II (i.e. exposure to police is measured by total visits for marijuana and home locations are estimated using the DPM). 95% confidence interval is obtained using the bootstrap method.

2.G Variable selection

In this section, we present the results from a least absolute shrinkage and selection operator (Lasso) regression. Lasso regressions estimate the effect of reducing variables to zero on the models prediction accuracy. While the interpretation of the coefficients is the same as standard logistic models, Lasso regression identify which covariates do not reduce prediction accuracy when omitted from the model. After crossvaligdation, we find that omitting both measures of arrests, proportion Hispanic, poverty rate, and population^2 significantly reduce prediction error. Other variables can be omitted. These results are presented in Table 2.G.1.

Table 2.G.1: Lasso Regressions of Protest Participation Using Social Activity Hub

	Model I Coef.	Model II Coef
Exposure to Policing†	0.06	0.05
Proportion Black	—	—
Proportion Hispanic	−2.12	−2.12
Proportion Female	—	—
Youth Bulge	—	—
Median Income	—	—
Poverty Rate	0.38	0.38
Unemployment Rate	—	—
Proportion HS Graduate	—	—
Per 1000 Population	—	—
Per 1000 Population^2	−0.01	0.01
Intercept	−0.64	−0.64

† Exposure to policing is per 100 arrests in Model I and per 100 visits in Model II.

Note that standard errors are not available from Lasso estimates.

2.H Influential Observations

One concern is that our results are driven by a few influential observations that bias our estimates. To assess whether this is the case, we perform two regression diagnostic tests. In the first, we estimate the DF Betas (D_{ij}) for our measure of exposure to policing. A DFBeta is defined as $D_{ij} = \beta_j - \beta_{j(-i)}$ for $i = 1, \ldots, n$ and $j = 0, 1, \ldots, k$ where β_i are for all observations and $\beta_{j(-i)}$ are those with the ith observation removed. Appendix Figure 2.H.1 shows that almost all observations lie along a single line, with no observations exceeding the critical value.

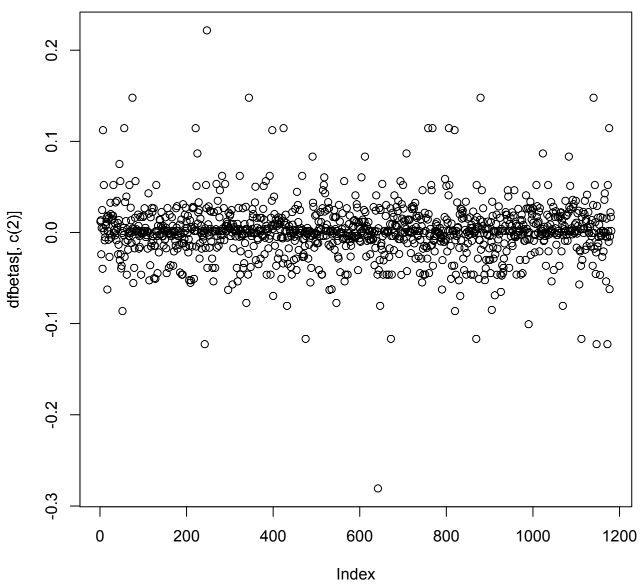

Figure 2.H.1: Figure of the DF Betas estimated for our first model.

To confirm this observation, we also estimate Cook's distance, which estimates an F-test for the hypothesis that $\beta_j = \beta_{j(i-)}$ for $j = 0, 1, \ldots, k$. Although there is no significance test for Cook's distance, the rule of thumb cutoff is $D_i > \frac{4}{n-k-1}$. The results from this estimation are presented in Appendix Figure 2.H.2. While three observations appear to exceed the critical value, omitting them from our analysis does not change any results.

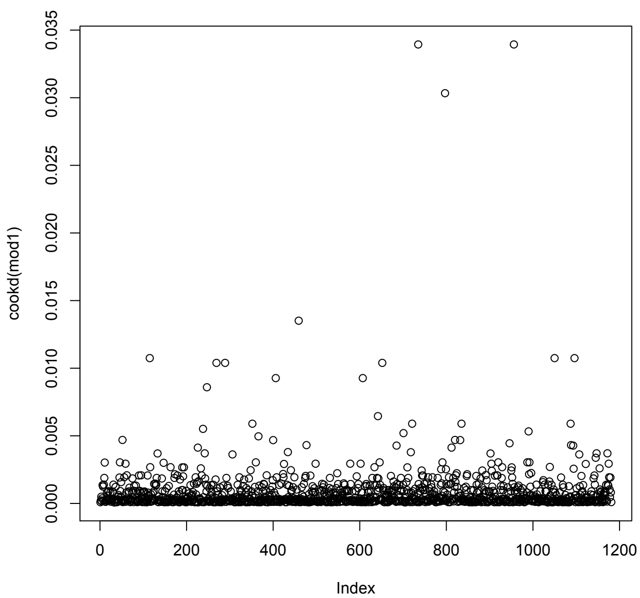

Figure 2.H.2: Figure of Cook's distance estimated for our first model.

2.H.1 Random Forest Results

Before continuing, it is important to contextualize the resulting permutation importance scores. Unlike regression coefficients, the size of the importance score does not have any meaning in and of itself so long as it exceeds zero. When a variable's permutation importance that is indistinguishable from zero suggests that randomly permuting that predictor variable has no impact on the model's out-of-sample error rare. It is possible to rank predictor variables in terms of their importance scores. As importance scores increase, it suggests that randomly permuting that variable increases the out-of-sample error rate. As the scores are not inherently meaningful, we present the ranking of the predictor variables from our model graphically. The results from this random forest are presented in Figure 2.H.3.

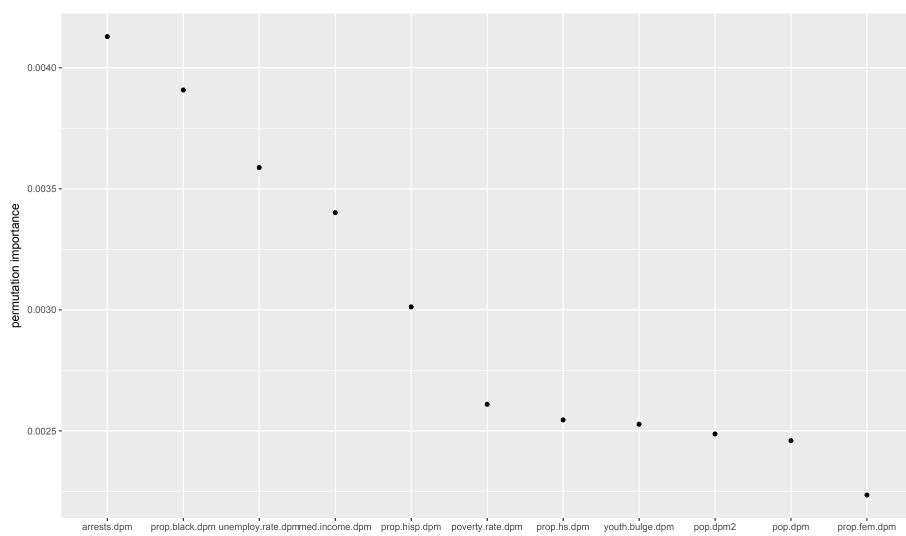

Figure 2.H.3: Permutation importance from random forest model.

As Figure 2.H.3 shows, our main measure of contact with the police (denoted as arrests.dpm) is the best predictor variable. This suggests that the results obtained in our logit are not the result of omitted interaction effects or non-linear effects.

3 Sunday, Bloody Sunday: Evidence from Northern Ireland for the Effect of Ethnic Diversity on Violence

Abstract

How does the spatial collocation of ethnic groups affect conflict intensity? Theory and empirical research suggest that diversity intensifies violence except when specific conditions exist to promote interethnic tolerance. We test this contention using uniquely high resolution data from the Troubles in Northern Ireland. Although Protestants and Catholics lived in close proximity for generations before the conflict, we show that they did not interact in ways that allowed meaningful social contact or social capital formation. Our empirical strategy corrects for measurement error and endogenous population dislocations, using the spatial distribution of Catholic and Protestant churches to construct an instrumental variable for ethnic diversity. We show that places with greater diversity experience more conflict-related deaths. Unadjusted estimates without a corrective empirical strategy underestimate diversity's conflict-intensifying effect by 50%.

3.1 Introduction

Theoretical and empirical scholarship have furthered our understanding of ethnic conflict, while leaving open some questions about its spatial dimension. Theory suggests that places with greater ethnic diversity – defined as the spatial collocation of ethnic groups – will experience greater violence (Kaufmann 1996; Posen 1993). Scholars have also outlined conditions under which diversity may promote peace and tolerance, provided that certain interethnic social or economic links exist (Bhavnani et al. 2014; Jha 2013; Pettigrew and Tropp 2006; Varshney 2001). Although it is assumed that collocation without interethnic linkages increases violence, this assumption has been inadequately tested in the literature. This is problematic because policy scholars frequently assert the desirability of interethnic contact and integration as peacebuilding measures in post-conflict societies such as Northern Ireland (Hughes et al. 2008; Richardson and Gallagher 2011; Tam et al. 2009).

Empirical researchers have sought to test this conjecture, particularly as improved geospatial data allow for more detailed analysis (Cederman and Gleditsch 2009). However, some data limitations persist. Most studies can only measure diversity at the level of administrative units at which data are available, such as districts or electoral wards.[1] This spatial resolution is reasonable when political processes like elections are the sole determinant of conflict onset and intensity (Dancygier 2010; Wilkinson 2006). If, however, the spatial dynamics of conflict are driven by interpersonal contact, large spatial units of analysis are inappropriate because they can obscure local segregation (Bhavnani et al. 2014; Field et al. 2008; Klašnja and Novta 2014). Some studies, moreover, use ethnic settlement data collected after conflict onset, introducing the possibility of severe measurement error. As populations rapidly segregate once conflict starts, it changes communities' ethnic composition and produces measurement error due to the difficulties of data collection during conflict (Weidmann and Salehyan 2013). Such dislocations endogenize future measurements of ethnic diversity to the violence, biasing the estimated relationship between the

[1] Wards are electoral subdivisions of Belfast.

two. These various data issues may explain the fragility of the empirical association between diversity and conflict (Ellingsen 2000; Fearon and Laitin 2003).

Using a novel empirical strategy, we demonstrate that diverse areas experience more intense violence during ethnic civil wars when salubrious economic or social interethnic linkages are absent. We improve on existing analyses using uniquely rich data from the Northern Ireland Troubles, a thirty-year ethnic conflict that resulted in over 3,000 deaths, including 1,617 in the capital city, Belfast (CAIN 2014). Unionist Protestants fought to maintain Northern Ireland within the United Kingdom, while nationalist Catholics attempted to secede. Catholics and Protestants lived in mixed neighborhoods for generations before the onset of violence in 1969. However, ethnographic evidence from Belfast shows that Catholics and Protestants did not build interethnic social networks and avoided face-to-face interaction (Boal 1969). In the absence of positive interethnic linkages, the spatial collocation of Protestants and Catholics exacerbated elite fragmentation and interethnic labor competition. In a case where social capital and social contact mechanisms cannot operate, theory predicts that ethnically diverse areas of Belfast should experience higher levels of violence.

Our empirical strategy benefits from the high spatial resolution of the 1971 UK census, which is available at 100-meter resolution in Belfast (Shuttleworth and Lloyd 2007). These cells, roughly the size of a city block, are an improvement over other studies' census tract or ward-level units since it is reasonable to assume that residents interact at least minimally (Balcells, Daniels, and Escribà-Folch 2016). This is important given that ethnic demography can often show large discontinuities because of ethnic sorting along focal boundaries.

To correct for attenuation resulting from measurement error and endogenous population sorting, we construct a novel instrumental variable based on the spatial distribution of Catholic and Protestant churches built before the Troubles. Ethnically mixed areas were more likely to experience conflict deaths compared to less diverse areas. Unadjusted estimates without an empirical strategy to correct for endogeneity and measurement error underestimate diversity's effect by 50%.

Our work makes several contributions. First, our empirics confirm the theoretical prediction that ethnically diverse areas experience more intense violence when social or economic interethnic linkages are absent. Second, we make a methodological contribution, highlighting the importance of research designs that account for endogenous population sorting and measurement error in ethnic settlement data. Third, we introduce new geolocated data to studies of conflict in Northern Ireland.

3.2 Literature on Diversity and Violence

Theoretical and empirical studies have examined the links between violence and the spatial collocation of ethnic groups. Collocation can influence the likelihood of conflict (Collier and Hoeffler 2004; Ellingsen 2000; Fearon and Laitin 2003). It may also influence the intensity of violence once conflict breaks out. However, conflict onset and conflict intensity may follow different logics (Kalyvas 2005). Our paper deals with conflict intensity and diversity's presumed effects at a local level. It also deals with "irregular" or "guerrilla" conflict, as opposed to "conventional" civil war with distinct front lines.[2]

Scholars identify several mechanisms through which diversity may intensify violence. The experience of collocation prior to conflict generates material and emotional grievances, which are unleashed by conflict onset. Diversity intensifies competition over public goods and cultural hegemony (Adida, Laitin, and Valfort 2015) and opens the possibility of competing homeland claims (Toft 2003). A larger population share gives aggrieved minorities more resources to mobilize, while making majority ethnic groups increasingly distrustful and hostile (Blalock 1967). Fear, hatred, and resentment motivate individuals and communities toward violence (Petersen 2002). Armed groups act on these grievances – and stoke them to encourage enlistment (Hardin

[2] Conventional civil wars involve "face-to-face confrontations between regular armies across clear frontlines," and are less common than irregular civil wars (Kalyvas 2005: 90-91).

1995). The build-up of fear, hatred, resentment, and rage heightens demand for violence against the ethnic "other," and armed groups strategically respond (Petersen 2002).

Other mechanisms relate to processes within a conflict. The collocation of ethnic groups creates a security dilemma, giving ethnic groups incentives to eliminate "the other" entirely from a given area (Posen 1993). The experience of violence hardens ethnic identities in diverse and formerly diverse zones (Acharya, Blackwell, and Sen 2016). This makes it easier to recruit combatants and harder to end a conflict without fully separating the warring groups (Brass 1997; Kaufmann 1996). Diversity also creates a zone of contested control, making it difficult for either side to distinguish combatants from civilians (Kalyvas 2006). Rebels may use indiscriminate violence to provoke an indiscriminate state response, driving coethnics toward the rebels' side (Wood 2010). If any of the above mechanisms, or any combination of them, operates, violence will be more intense in ethnically diverse areas.

A separate body of scholarship identifies mechanisms that moderate diversity's presumed negative effects. Social contact theory argues that increased interaction improves intergroup trust and understanding (Allport 1954). Socially generated mediating factors – knowledge about ethnic others, decreased anxiety about intergroup contact, and increased empathy – should attenuate violence in diverse areas where ethnic groups interact positively. There is empirical support for these predictions in ethnic conflict settings where politico-economic conditions do not preclude meaningful social contact (Hewstone and Swart 2011; Lemmer and Wagner 2015; Pettigrew and Tropp 2008). Social capital theory gives similar predictions. When groups interact in shared social organizations that cut across ethnic lines, "bridging social capital" enables elites to manage tensions and resolve disputes (Putnam 2000: 32, Varshney 2001). Economic interactions create further incentives to learn about ethnic others, rather than believing stereotypes and scurrilous rumors (Glaeser 2005). When ethnic groups occupy distinct but interdependent economic sectors, labor complementarity encourages elites to develop violence-dampening institutions (Jha 2013). If social and economic conditions allow these conflict-reducing mechanisms to operate, we should expect ethnically diverse areas to be less violent, rather than more violent.

3.3 Data Limitations

Empirical studies attempting to test these propositions suffer from three important, yet often overlooked, data limitations: the use of geographically large units of analysis; endogeneity in data collected after conflict onset; and potentially severe measurement error. We consider these problems in turn.

3.3.1 Unit of Analysis

Ecological inference problems arise when using excessively large geographic units. For example, the observed relationship between diversity and conflict outbreak depends on the unit of analysis and the operationalization of diversity (Cunningham and Weidmann 2010). Demographic diversity in a given area does not imply a corresponding level of interethnic contact or social integration (Hewstone and Swart 2011). A "micro-ecology of segregation" may persist (Dixon et al. 2008), with different groups managing to self-segregate, even within the same school cafeteria (Clack, Dixon, and Tredoux 2005).

Recent studies using subnational data are a welcome step forward, although administrative-level units may still lack the resolution to assess theories. For instance, Klašnja and Novta (2014) predict that in highly polarized municipal districts, diversity will promote the spread of conflict. Successful uprising in one municipality increases rebels' probability of success in neighboring municipalities. The authors test their predictions using census data from India and Bosnia, where administrative regions are broken down into tehsils and settlements, respectively. These units are adequate for testing conflict contagion theories, but not theories with social contact or social capital mechanisms. As the authors note, tehsils "contain multiple villages and a few towns" and Bosnian settlements have a median population of 1,400; "certain blocks within a city or even a whole village might be either Hindu or Muslim," rendering face-to-face interaction unlikely (*Ibid,* 10, 15).

Balcells, Daniels, and Escribà-Folch (2016) also demonstrate the importance of obtaining sufficiently high-resolution data. The authors predict that in post-civil war settings, "low-intensity conflict" will be highest in polarized areas where ethnic groups are locally segregated. The authors use data from the 584 census wards of Northern Ireland – the smallest geographic unit for which local police report riot statistics. Unfortunately, ward boundaries do not correspond with ethnic settlement patterns. A polarized ward may contain several segregated neighborhoods, or it may be diverse throughout.

3.3.2 Endogeneity and Measurement Error

One immediate consequence of civil war is population movement away from violent areas. Seeking safety in numbers, residents flee contested areas and relocate to areas controlled by their ethnic group. In Belfast, the August 1969 onset of violence was responsible for at least eight deaths; the destruction of 150 Catholic homes; and the dislocation of 2,069 families, mostly from diverse areas (CAIN 2014; Coogan 1993; Griffiths 1971). Within a matter of weeks, 75% of Belfast residents retrenched to areas that were 90% Protestant or 90% Catholic (Boal 1982; Irvine 1991). Census data collected after conflict onset understate the degree and spatial extent of pre-conflict diversity, biasing the estimated relationship between diversity and violence toward zero.

Surveys taken during a conflict are also subject to measurement error. In Belfast, many residents resisted the 1971 census, fearing that their answers would be shared with police (Gallagher 1971: 15). Resistance was particularly strong among Catholics: 26 priests urged Catholics not to complete the forms (O'Connor 1971: 15) and the IRA set fire to the Ministry of Finance where completed forms were kept (Fermanagh Herald 1971; Sunday Independent 1971: 1). The resulting measurement error is correlated with diversity, generating biased estimates in the absence of a corrective empirical strategy.

Measurement error also arises in pre-conflict data. Kopstein and Wittenberg (2011: 260) find that areas of Poland with greater "intercommunal polarization between Jews and the titular majority" experienced more pogroms in 1941. Data from the 1930s and 1940s are unavailable, so the authors use the best available data: a 1921 census, which suffers from systematic over-counting of Poles (*Ibid,* 266). Low-violence zones that appear diverse may in fact have been majority Jewish, biasing the results.

Kasara (2016), likewise, finds that diverse areas experience less violence, based on data from Kenya's 2007/8 post-election crisis. This study instruments for diversity using the number of farm subdivisions cross-cutting each administrative unit. However, the farm subdivisions were measured in 1964, raising questions about whether some boundaries are endogenous to decolonization, land reform, or the 1952-1960 Mau Mau uprising. Data limitations also force Kasara to use fires and population displacement as proxies for violence. Laudable for its sophisticated use of the available data, the study nonetheless raises ecological inference concerns: its administrative unit has "an average population of 9,000 and a median area of 22 square miles" (*Ibid,* 12).

3.4 The Troubles In Northern Ireland

3.4.1 The Troubles as a Theory Testing Case

We adopt the following definition of ethnicity from Chandra (2012: 9): "a subset of categories in which descent-based attributes are necessary for membership…including categories based on the region, religion, sect…of one's parents or ancestors." The Northern Ireland Troubles were an ethnic conflict between Catholic nationalists and Protestant unionists, each side with its own distinct ancestry, traditions, and identity (Hancock 1998).

Northern Irish Catholics trace their ancestry to Ireland's ancient *Gaeilge*-speaking inhabitants, who adopted Christianity in the 5th Century (Hughes 1966). Protestants trace their ancestry to the Scots and English who colonized "Ulster" (Northern Ireland) in the 17th Century. The "Ulster Plantation" was part of England's effort to dispossess Catholic landlords and consolidate royal control of Ireland. It also established Northern Ireland's enduring ethnic cleavage: nationalist Catholics versus pro-British Protestants.

Ethnicities in Northern Ireland are disjoint; one cannot be simultaneously Catholic and Protestant. One is born into one ethnicity or the other. In mixed marriages (accounting for less than 2% of all marriages) households typically adopt either Catholicism or Protestantism (Robinson 1992). Ethnic identification is reinforced by attendance at church and politically charged parades, such as the Orange Marches (celebrating William of Orange's Protestant conquest of Ireland) and Catholic Saint Patrick's Day parades. Other markers communicate an individual's ethnicity in social settings. Names such as William and Andrew are markers of Protestant heritage, carrying the legacy of William of Orange and Andrew, patron Saint of Scotland. Names such as Seán and Mairéad are Irish in origin and are common only among Catholics.

Political behavior in Northern Ireland follows ethnic lines. In 1968, prior to the Troubles, 75% of Catholics identified themselves as "Irish," while 25% identified as "British" or "Ulster." Just 20% of Protestants identified as Irish (Moxon-Browne 1991). Although only a minority of Northern Ireland's residents participated in violence during the Troubles, participants adhered to sectarian lines. The Irish Republican Army (IRA) drew its forces from Catholic neighborhoods. IRA violence against the state predominantly killed Protestants, because police and local military forces recruited almost exclusively from the Protestant population. Loyalist paramilitaries reacted to IRA attacks by calling Protestants to resist "the forces of Romanism" (UVF 1971).

3.4.2 Social contact and social capital in Northern Ireland

We use historical data to assess the conditions in our study area of Belfast. The consensus among observers is that despite high levels of diversity in the city, Belfast shows little evidence of social contact or social capital-building between Protestants and Catholics. In terms of economic opportunities, moreover, Catholics and Protestants were highly horizontally unequal (Cederman, Weidmann, and Gleditsch 2011). Catholics were excluded from economic advancement and formed separate business institutions.

Belfast's labor markets created interethnic competition rather than complementarity. The 19th Century migration of Catholic laborers from the countryside created competition for jobs in Belfast's industrializing economy. Unskilled and low-skilled Catholic and Protestant laborers were interchangeable, unlike the Hindu craftsmen and Muslim merchants studied by Jha (2013). Linen mill owners exploited competition to keep wages down (Hepburn 1996). The growth of the Catholic population and the perceived presence of a "reservoir" of additional Catholics, ready to "invade" Protestants' neighborhoods and take their jobs created an environment of perceived scarcity and opposition (Boal 1982; O'Hearn 1983).

Competition in high-skilled sectors was exacerbated by anti-Catholic discrimination (Lynch 1998). Protestant shipyard owners imported skilled English and Scottish laborers, who shut Catholics out via a hereditary apprenticeship system and violence against Catholics who managed to gain apprenticeships. Shipyard owners "had no particular interest in widening the market beyond a Protestant pool that was already larger than necessary;" employers tolerated discrimination in the name of "workplace harmony" (Hepburn 1996: 26).

Job discrimination was compounded by a segregated education system in which Protestant schooling was perceived to be of higher quality. Employers, meanwhile, were located in Protestant areas without public transit links to Catholic neighborhoods. These factors segregated workplaces so that Protestants dominated well-paid skilled positions. Catholics were confined to low-skilled sectors and suffered disproportionate unemployment – 14% versus a Protestant average of 6% (O'Hearn 1983, Rowthorn 1981:6).

Socially, Catholics and Protestants did not interact within the same voluntary organizations as envisioned by social capital theorists. Protestants shut Catholics out of public bodies representing businessmen and civil servants; Catholics established parallel business organizations, sports leagues, and newspapers (Hepburn 1996). Fraternal organizations such as the Orange Order and the Ancient Order of Hibernians were designed to exclude outgroups, reinforcing Protestant and Catholic ingroup cohesion.

Catholic and Protestant churches encouraged the further segregation of social spheres. The Catholic Church insisted that parishioners be instructed by Catholic teachers (Irvine 1991). A hyper-segregated Catholic school system developed alongside a *de facto* Protestant public system (Knox 1973). As of 1960, 98% of Catholic children attended Catholic primary schools (Barritt and Carter 1962: 77). In 1977, 71% of Northern Ireland's children attended schools with zero children of the other ethnic group (Murray 1985: 31). Just 3% attended schools with 5% or more of the other group (Darby 1978: 217).

Catholic and Protestant religious leaders encouraged their adult flocks to build "two almost mutually exclusive sets of [social] networks" (Boal 1969: 359). They established competing temperance societies and banned Easter celebrations at which Protestants and Catholics previously mixed (Hepburn 1996). Proscribing "British" games such as rugby in Catholic schools further reduced the likelihood of Catholics and Protestants mixing at sporting events in adulthood (Barritt and Carter 1962: 148).

Ethnic retrenchment spilled over into street life as well. Boal (1969) finds that Protestants and Catholics actively avoid interactions that would increase their visibility to the outgroup. In mixed streets and neighborhoods, they rarely visit one another's homes, avoid shopping at the same stores, and strategically select bus stops to avoid walking through areas dominated by the other group. As a result of their self-segregation, Protestants and Catholics may have different names for neighborhoods they jointly inhabit.

3.5 Sources of Data

Among all the world's ethnic conflicts, Northern Ireland's is exceptionally data rich. Government and news reporting minimize the amount of missing data on key variables such as the number and locations of conflict deaths. When we do encounter missing data, qualitative evidence allows us to validate our interpolated measures. Appendix Table 3.A.1 presents the summary statistics for all variables used, including the diversity variables constructed from the raw census and interpolated values.

3.5.1 Population and Diversity

Northern Ireland conducted census surveys in 1962 and 1971, both within a decade of the Troubles' onset. Although both censuses asked respondents about their faith, the 1962 census results were aggregated and reported for the 15 electoral wards of Belfast at the time of the census. Disaggregated data are unavailable, and using the 1962 data would yield only 15 observations. More problematic still, the exact ward boundaries are unknown. The city council reorganized its ward boundaries several times between 1950 and 1973, and the census does not specify which boundaries were used.

The 1971 census is available at a 100-meter grid square resolution for all urban areas within the United Kingdom, including Belfast (Shuttleworth and Lloyd 2007). We use these data within Belfast city boundaries as our unit of analysis. These data offer a much higher resolution than ward- or municipal-level studies (Balcells, Daniels, and Escribà-Folch 2016; Klašnja and Novta 2014). The high resolution allows for measuring diversity and its effects in a given city block. Appendix Figure A1 shows the distribution of the population in Belfast in 1971 using these data. We limit our analysis to Belfast because the city's rapid industrialization produced a workforce that was significantly more diverse and integrated than populations elsewhere (Hechter 1977). Because the precise 1971 census boundaries of Belfast are unknown, we use the 2005 urban extent defined by the Northern Ireland Statistics and Research Agency (NISRA 2005).

We code our measure of diversity by using the 1971 census data to estimate the number of Catholics and Protestants in each grid cell. As noted in Section 3.2, the raw census data are compromised by the endogenous population dislocations of 1969-1971. They also suffer from measurement error: widespread resistance (particularly among Catholics) forced census enumerators to declare the survey's faith questions optional (Gallagher 1971: 15).[3] As a result, 34% of cells in our dataset (2,807 of 8,258) reported population data but not faith. In cells without religion data, we interpolate the percent of Catholic residents using ordinary Kriging.[4] The procedure fits a semivariogram function relating the autocorrelation of the data points as a function of distance, and then interpolates at each location using the nearest 12 census data points. We use the interpolated values for the Catholic percentage of the population to construct a measure of diversity for each 100-meter cell (indexed by i) representing the extent of the 1971 census. Appendix Figure A2 visualizes the interpolation in the highly contested Divis neighborhood.

$$EthnicDiversity_i = 1 - \left| \frac{(CatholicPopulation_i - ProtestantPopulation_i)}{(CatholicPopulation_i + ProtestantPopulation_i)} \right| \tag{3.1}$$

The variable approaches a value of 1 when the Catholic percentage of the population approaches 50% (i.e. when each grid cell is more polarized) and approaches zero in entirely Catholic or Protestant cells.[5]

3.5.2 Church Density

We construct an instrument for ethnic neighborhood composition using the spatial distribution of Catholic and Protestant churches built before 1969. We rely on multiple sources to find

[3] The census had 27 questions divided into two sections: a demographics section and a supplemental section containing questions about faith and socioeconomic variables (McKeown 1971).

[4] Appendix Table 3.N.1 shows that our results are robust to limiting the sample to areas where data on religion were available. This shows that our results are not driven by locations for which religious composition was interpolated.

[5] Results are all robust to using the Herfindahl index, which has a correlation of -0.97 with our diversity measure.

the locations of these churches. We first compiled a list of churches that currently hold religious services from the websites of Ireland's five main Christian denominations: Catholics (39 churches in or within three kilometers of Belfast), Baptists (25), Presbyterians (83), Methodists (42), and the Church of Ireland (75).[6] These websites are designed to help parishioners locate religious services and do not provide information about churches that have closed or relocated. Failing to account for closed churches could bias our estimates by understating pre-conflict diversity. We consulted numerous historical sources and emailed the parish offices of all denominations to identify derelict or demolished churches that were open before the onset of the Troubles. We consulted parish histories and other secondary sources to confirm that the churches in our database predate August 1969.

We geolocated each church using street addresses from church websites and historic street directories. We used Google Maps' satellite imagery and public API to confirm each structure's latitude and longitude.[7] Figure 3.1 shows the locations of these churches.

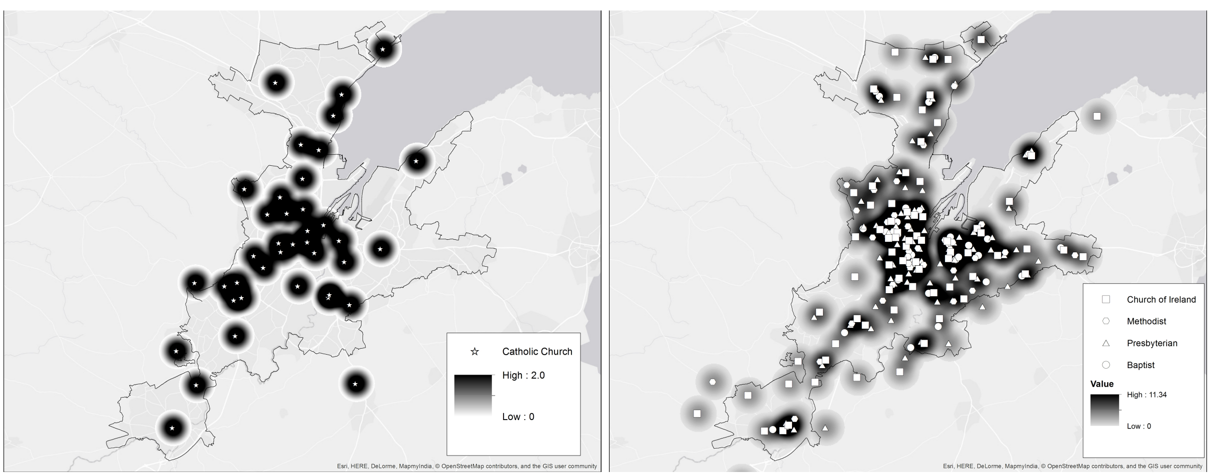

Figure 3.1: **Left:** Catholic Churches, location and kernel density with 1 kilometer bandwidth. **Right:** Protestant Churches, location and kernel density with 1 kilometer bandwidth.

[6] These denominations represent 95% of churchgoers in Northern Ireland, according to the 2011 census.
[7] See replication data notes for a list of sources consulted.

3.5.3 Deaths

Our data on the location of violent deaths come from the Conflict Archive on the Internet project at the University of Ulster (CAIN 2014). Researchers geolocated the 3,379 civilian and combatant deaths during the Troubles to the exact street address. The data are available as a Google Maps layer, from which we scraped the public XML file.[8]

We match each violent death to the nearest 1971 census grid cell and aggregate the deaths in that cell. (The median distance from the death to the center of a cell is 71 meters.) 90% of the 8,258 cells in Belfast experienced no deaths. The remaining cells experienced between 1 and 23 deaths over the 1969-2001 span of the data. Figure 3.2 shows a density plot of deaths in our dataset.

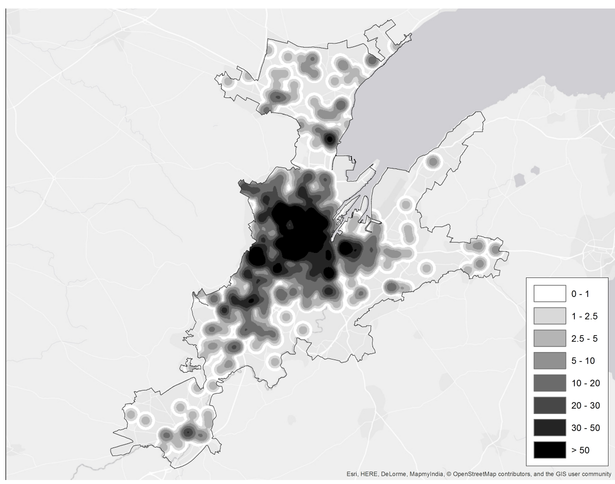

Figure 3.2: Plot of 500 meter bandwidth kernel density of violent incidents. Darker shades represent more violent areas.

[8] Original data available at `http://cain.ulst.ac.uk/victims/gis/googlemaps/victims.html`. CAIN is not the only source of geospatial data from the Troubles. The Northern Ireland Research Initiative (NIRI) uses machine learning to estimate the location of violent events (Loyle, Sullivan, and Davenport 2014). However, the NIRI dataset contains significant coding errors – for instance, assigning over 80 deaths to the same point in central Belfast and over 30 to a field west of Belfast. These locations are not supported by qualitative accounts of the conflict (McKittrick et al. 2001).

3.6 Estimation Strategy

We use the location of churches predating the 1969 onset of the Troubles to construct an instrument for ethnic diversity as measured by the 1971 census. We estimate kernel densities stretching out from each church, generating 25-meter grids of each denomination's church density per square kilometer in Belfast. We have no prior on the range over which a church influences parishioners' location decisions, so we construct kernels stretching 500 meters, 1 kilometer, 1.5 kilometers, 2 kilometers, and 3 kilometers from the church. We report results at varying kernel bandwidths.[9] We construct the instrument for the religious diversity of the population by creating a variable for church diversity:

$$ChurchDiversity_i = 1 - \left| \frac{(CatholicChurchDensity_i - ProtestantChurchDensity_i)}{(CatholicChurchDensity_i + ProtestantChurchDensity_i)} \right| \tag{3.2}$$

Cells with a value close to 1 have a similar density of Protestant and Catholic churches, indicating a highly diverse community. Cells with a value near zero have significantly more Catholic or more Protestant churches, indicating low diversity. We construct the instrument at varying sizes of the kernel densities around each church, from 500 meters to 3 kilometers. Cells with neither Catholic nor Protestant churches within the kernel bandwidth have no value for the instrument and are not included in regressions at that kernel size. Figure 3.1 plots the results of the church diversity instrument for the 1 kilometer bandwidth.

[9] We use identical bandwidths for Catholic and Protestant churches. We also show in the Appendix that the results are robust to allowing different bandwidths across denominations. Given the history of Belfast's urbanization (see Section 6.1) and the importance of building churches within walking distance of parishioners, we have no reason to suspect that parishioners of either denomination would live closer to their churches or be more likely to relocate away from their parishes.

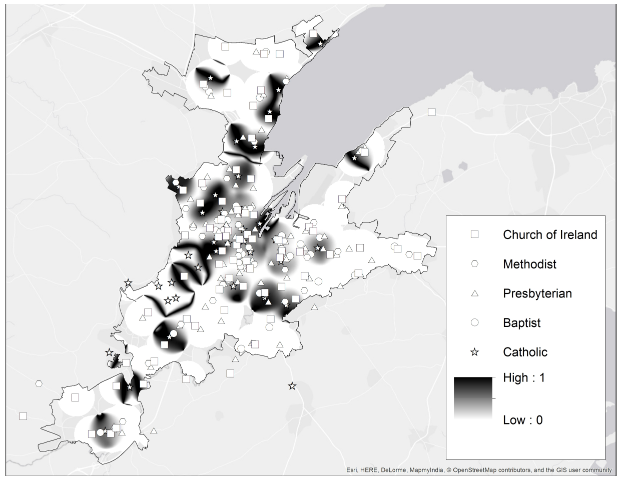

Figure 3.1: Diversity of churches constructed from Catholic and Protestant kernel densities in Belfast. Darker areas are more mixed in church density.

3.6.1 IV Assumptions

First Stage Relationship

An instrumental variable (IV) estimation strategy requires that church placement correlate with population concentrations of Catholics and Protestants prior to the conflict. This relationship is well established, based on the history of settlement in Belfast and the late introduction of motorized transportation.

Belfast's burgeoning linen and shipbuilding industries attracted laborers from the countryside and the population grew from 53,000 in 1831 to over 400,000 a century later (NISRA 2005). The rapid growth is illustrated in Appendix Figure A3. Without a public transit system, workers needed to live within walking distance of the mills. Mill owners commissioned hundreds of low cost family homes adjacent to their facilities (Bardon 1982). This housing attracted thousands of new laborers to Belfast's industrial neighborhoods.

Local religious orders struggled to keep up with the demand for their services. The challenge was not only to have enough pews for Belfast's new inhabitants, but to have those pews within walking distance of parishioners' homes. The issue was particularly acute for Catholics, who experienced the greatest growth – from roughly 2,000 (10% of Belfast's population) in 1800 to 40,000 (one third) in 1861 (Irvine 1991). Prior to 1850, the Catholic Church had only three houses of worship in Belfast, none within easy walking distance of the fastest growing center of Catholic population, the industrial Pound and Falls area. Catholic leaders erected a cathedral in the Pound and Falls, plus fifteen parish churches in the 1870s and 1880s. They selected the sites strategically to minimize the distance between churches and mill workers' homes (McGee 2013). The industrial population boom, though less pronounced among Protestants, similarly obliged Protestant denominations to build new churches in rapidly growing areas.

By churches' own accounts, the relationship between population growth and church construction remained consistent through the 20th Century. For instance, the website of Saint Agnes's church, in the Andersonstown neighborhood, explains that "the sudden explosion in house building after the [Second World] war…made the erection of a new church an urgent necessity."[10]

Independence

The independence criterion requires that the location of churches and the local collocation of Protestants and Catholics not be driven by a third, confounding factor. This situation might arise if legal discrimination prevented one or more religious denominations from settling or building churches in a particular area. The 1829 repeal of Britain's Penal Laws removed such restrictions, enabling the 19th Century building boom among non-Church of Ireland denominations. A violation of the independence criterion might also occur if episodes of ethnic violence prior to the Troubles influenced settlement patterns and the location of churches. Belfast did suffer periodic rioting and more substantial violence during the 1918-21 Irish revolution and 1922-23 civil war. However, these events did not disturb settlement patterns or church building.

When Catholics comprised a small minority of Belfast's population, Catholics and Protestants got along well. Protestants contributed much of the funding to build Belfast's first and second

[10] See `http://www.stagnesbelfast.com/?page_id=50`.

Catholic churches, in 1784 and 1815 (Irvine 1991). Interethnic relations soured as the Catholic population grew, and a pattern of periodic rioting emerged. Belfast experiences small riots every summer, when Orange Marches pass provocatively through Catholic neighborhoods. "Marching season" riots are typically limited to rock throwing and street scuffles, but at times, the ritualized violence has gotten out of hand. Irvine (1991: 55-6) observes a pattern of "eruptions... about every ten years." The 1886 riots were the worst, causing 31 deaths; a bout of rioting in 1935 killed 11 people. The organized violence of the Irish Revolution and Irish Civil War produced worse results, including roughly 465 fatalities in Belfast.

These episodes were far less serious than the Troubles, which killed 1,617 in Belfast alone. Moreover, the riots of the 19th and early 20th Centuries did not disturb the long term settlement and residential mixing trends established during industrialization. According to Boal (1982: 268), "mixed and segregated residential areas have existed in Belfast from the mid-nineteenth century onwards ... [T]he patterns observable in the 1970s are a continuation of something long established. Major outbursts of conflict have, from time to time, sharpened segregation, followed by some relaxation and an expansion of residential mixing." In fact, "some of the mixed residential areas have been locationally stable for long periods of time, perhaps contracting during outbursts of open conflict and then growing again during less stressful periods." This continuity is due in part to "the presence ... of ethnic institutions (particularly schools and churches) affiliated with both groups" (*Ibid.* 274). Individual families relocated within Belfast or migrated out of the city following episodes of rioting, but subsequent inmigration by coethnics (by cross-town moves or migration from rural areas) repopulated damaged areas with an ethnic mix similar to the one preceding the riots. Like the corresponding mix of physical churches, Belfast's pre-Troubles "ethnic residential mixing ... displayed considerable locational stability" (*Ibid.* 268).

A robustness check in Appendix Table 3.E.1 shows that controlling for socioeconomic status does not alter our estimates of diversity's effect. If socioeconomic status were the true driver of violence, and if it were highly correlated with church density and ethnic diversity, our instrument would improve the measurement of diversity's association with violence, but it would not provide a valid causal estimate. Qualitative accounts give us no reason to suspect an association between socioeconomic status and church density. Moreover, the correlation between diversity and socioeconomic status (-0.02) is quite low. Controlling for socioeconomic status does not substantively alter our IV estimates of diversity's effect. These results help to validate our independence assumption, while bolstering the argument for a causal relationship between diversity and violence in our study context.

Exclusion Restriction

Our exclusion restriction assumes that the diversity of Protestant and Catholic churches correlates with violence only through a relation to neighborhood demographics. This assumption would be violated if, for example, church pastors were more likely to espouse or denounce violence based on their church's proximity to churches of the other faith – *and* congregations acted on their pastors' wishes.

Our first piece of evidence against this argument is that Catholic and Protestant clergy condemned violence throughout the Troubles, with no discernible effect on violence levels anywhere. Paramilitary killings violated the Christian prohibition against murder, putting IRA and loyalist paramilitaries at odds with religious teachings. Catholics joining the IRA did so "against the wishes of the church"(Bell 1997: 463). Bishop Cathal Daly of Derry branded the IRA "an evil and barbaric organisation" and Pope John Paul II visited Northern Ireland in 1979 to demand an end to the violence. The IRA ignored these exhortations, accusing the Church of "politics and deceit" and betraying "the Christian values of truth and justice" (Moloney 2002: 232). Local clergy were also unable to affect violence levels. Priests were reluctant to open "political and other fissures within their congregations" by taking on paramilitaries directly. They appealed for peace while continuing to minister to their parishioners, including militants (Brewer and Teeney 2011: 181).

Although church leaders had no appreciable effect (positive or negative) on the killing, their properties were effectively off limits to paramilitary organizing. This is important for our exclusion restriction, which might be violated if paramilitaries attracted violence to churches by hiding or storing weapons inside. Throughout the conflict, paramilitaries held meetings and maintained their weapons caches in abandoned buildings and the homes of local families (Moloney 2010: 66). Targeting churches made little sense militarily, and the IRA regarded such attacks as "beyond the pale" (Brewer and Teeney 2011: 182). On the rare occasions when loyalists attacked churches, these attacks were unrelated to the presence or defense of Protestant churches nearby. The Siege of Saint Matthew's on June 27th, 1970, is the most notable example. IRA and loyalist gunmen exchanged fire near the church for several hours, killing three (Moloney 2002). Yet the shootings were only one incident on a night of city-wide rioting that killed six, and the involvement of the church was incidental: nationalist youths provoked a riot by throwing rocks at a loyalist parade; IRA gunmen took up sniping positions to defend the neighborhood against retaliation.

In sum, church leaders produced no systematic impact on violence levels and church buildings were not attractive targets for paramilitaries. Moreover, our instrumental variable is the diversity (not density) of churches. A violation of the exclusion restriction would require that church *diversity* affect levels of violence. We see no evidence of this.

As a robustness check against church diversity increasing (or decreasing) violent collective action via mechanisms other than demographics, Appendix Table 3.L.1 shows that our estimates of diversity's effect are substantively and statistically consistent after excluding cells in the immediate vicinity of churches.

3.7 Specifications and Results

3.7.1 Unadjusted Estimates

We first explore the association between diversity (as measured in the 1971 census) and the location of deaths during the Troubles using ordinary least squares (OLS) regressions. The regressions are instructive in showing the raw association, but they are unadjusted in that they do not account for pre-census population dislocations or measurement error in the census. One would naturally expect densely populated places to experience more deaths, potentially generating spurious correlation. We therefore control for population density using linear and quadratic terms. Finally, spatial autocorrelation between variables such as ethnic diversity, church density, and fatality counts raises challenges to statistical inference. We follow convention by nonparametrically estimating the variance-covariance matrix, allowing for spatial autocorrelation up to 500 meters (five grid cells). These standard errors are spatially heteroskedasticity autocorrelation consistent and are reported in square brackets (Conley 1999). They can be understood as a spatial analogue to Newey-West standard errors, allowing for arbitrary autocorrelation decaying up to a specified maximum distance.

$$Violence_i = \beta_0 + \beta_1 EthnicDiversity_i + \beta_2 Population_i + \beta_3 Population_i^2 + \varepsilon_i \tag{3.3}$$

Because the number of deaths in a cell has a lower bound of zero, we follow the unadjusted (and misspecified) OLS regression with a Tobit regression, specifying the dependent variable's lower bound. A third model employs a Poisson specification, given that the dependent variable is a count of deaths. Although these Tobit and Poisson models still fail to account for endogeneity and measurement error, they are better suited to our dependent variable.

We note that the underlying process determining where deaths occur is not necessarily the same as the process determining how many deaths occur in violent places. We estimate additional models to explore what may be thought of as the extensive and intensive margins of violence: which cells experience violent deaths (extensive) and the number of deaths in locations that experienced some deaths (intensive). We estimate a logit regression where the dependent variable is whether the grid cell experienced any deaths. We estimate an OLS regression on cells with nonzero deaths to determine whether the 1971 diversity measure is correlated with the number of deaths in these cells.

The results are shown in Table 3.1. In column (i) the association between our diversity measure and number of conflict deaths has a coefficient of 0.40. Cells with the maximum diversity score (1) have on average 0.4 more deaths than cells with the minimum score (0). A standard deviation increase in diversity (0.28) is associated with 0.11 more deaths, roughly half the standard deviation of deaths (0.20) in the data. The relationship is significant at the 99% confidence level. The spatial autocorrelation-adjusted standard error on the diversity variable (0.07) is almost twice as large as the unadjusted standard error (0.04), underscoring the importance of correctly modeling the spatial structure of the data. The correlation between number of deaths and population is strongly positive and quadratic: at the mean value of log population (3.67) a 1% population increase is associated with 0.13 more deaths.

Column (ii) presents the results of the Tobit regression, accounting for the dependent variable's zero lower bound. The coefficient on diversity increases almost tenfold to 3.45, illustrating the bias from misspecification using simple OLS. Diversity is still significant at the 99% confidence level, using regular standard errors because Conley spatial standard errors can only be implemented in OLS and logit specifications. Column (iii) implements the Poisson regression designed for count data. The coefficient on diversity is 1.64, significant at the 99% confidence level. The association between diversity and the number of deaths is positive and statistically robust to Tobit and Poisson specifications.

Column (iv) limits the sample to cells that experienced at least one death. This reduces the sample from 8,258 to 785, and allows us to explore the association between ethnic diversity

Table 3.1: Unadjusted Regressions of Violence and Diversity

	(i) OLS	(ii) Tobit	(iii) Poisson	(iv) OLS on Intensive Margin	(v) Logit on Exten- sive Margin
Ethnic Diversity	0.40***	3.45***	1.64***	1.05***	1.46***
	(0.04)	(0.31)	(0.08)	(0.27)	(0.14)
	[0.07]	-	-	[0.33]	[0.19]
ln(Population)	-0.36***	-2.81***	-1.05***	-0.45**	-1.30***
	(0.03)	(0.27)	(0.07)	(0.25)	(0.12)
	[0.05]	-	-	[0.22]	[0.13]
ln(Population)^2	0.07***	0.54***	0.20***	0.07***	0.25***
	(0.005)	(0.04)	(0.01)	(0.03)	(0.02)
	[0.010]	-	-	[0.03]	[0.02]
N	8,258	8,258	8,258	785	8,258
R^2	0.05	-	-	0.03	-
Pseudo R^2	-	0.07	0.12	-	0.10

Standard errors reported in (); spatial autocorrelation-adjusted errors reported in [].
Inference uses spatial autocorrelation-adjusted standard errors when available.
$^{**}p<0.05$; $^{***}p<0.01$
Regressions include a constant, not shown.

and the intensity of conflict in places that experienced some deaths. The coefficient on the diversity variable is 1.05, significant at the 99% level whether or not standard errors are adjusted for spatial autocorrelation. Among cells experiencing violence (their average number of deaths is 2.06 with a standard deviation of 2.18) a maximally diverse cell has on average 1.05 more deaths than a homogeneous cell. Column (v) shows the logit model, exploring the extensive margin: where deaths occurred, regardless of number. The coefficient on the diversity variable is positive and significant at the 99% level after flexibly controlling for population size. Note that we expect these unadjusted regressions to underestimate the true impact of diversity on violent deaths; comparing the IV estimates to the coefficients in Table 1 highlights the importance of an appropriate econometric strategy given the estimation challenges researchers face.

3.7.2 Instrumental Variable Regressions

Our IV specification corrects for the likely attenuation bias resulting from pre-1971 population dislocations and measurement error in the census. The first and second stage specifications are shown in equations (4) and (5), respectively, with the diversity of churches instrumenting for ethnic diversity in each grid cell.

$$EthnicDiversity_i = \gamma_0 + \gamma_1 ChurchDiversity_i + \gamma_2 Population_i + \gamma_3 Population_i^2 + u_i \tag{3.4}$$

$$Violence_i = \delta_0 + \delta_1 \widehat{EthnicDiversity}_i + \delta_2 Population_i + \delta_3 Population_i^2 + v_i \tag{3.5}$$

Table 3.2 shows the estimation results. To test the sensitivity of our results to kernel size, each of our tables reports five kernel bandwidths, ranging from 500 meters to 3 kilometers. The top panel reports first stage regressions at different bandwidths. In all cases, the instrumental variable (diversity of Catholic and Protestant church densities at each location) is positively associated with the ethnic diversity variable, even after controlling for population size. The relationship is significant at the 99% level with conventional and spatially-adjusted standard errors, and the instrument is strong (Staiger and Stock 1997).[11] Note that the sample size increases as the kernel bandwidth around churches increases. This is because cells beyond the kernels of all churches have no value for the church diversity instrument and are excluded from the regression. Only with the 3 kilometer bandwidth does the sample reach the same size as in Table 1. The significance and magnitude of the association between the church diversity instrument and ethnic diversity remain consistent across bandwidths.

The second stage of the estimation is shown in the bottom panel, using a Tobit specification with a lower bound of zero for the dependent variable. The instrumented value of ethnic diversity has a strong positive association with the number of victims in a grid cell. The coefficient is significant across bandwidths. Its magnitude, ranging from 4.62 to 8.71, indicates that a grid cell one standard deviation more diverse (0.25 in our diversity measure) experienced 1.2-2.2 more deaths. This is more than the standard deviation of the number of deaths (0.9).

Note that the IV Tobit coefficients are always larger than the unadjusted Tobit coefficient (3.45) in column (ii) of Table 1. A Hausman specification test comparing the coefficient in the unadjusted Tobit regression with any of the IV Tobit coefficients in Table 2 rejects the null hypothesis that the differences in coefficients are not systematic ($p = 0.01$). The larger magnitude of the coefficient on ethnic diversity in the IV specification is consistent with our expectations regarding the effects of measurement error and pre-census population sorting. Indeed, the coefficient on diversity in the IV Tobit specifications is roughly double the magnitude estimated in the unadjusted Tobit specification. This underlines the importance of accounting for endogeneity and measurement error when measuring ethnic diversity.

The results in Table 3.2 show that there is a strong positive association between the spatial collocation of ethnic groups and the intensity of conflict violence. In the appendix, we report alternative specifications as robustness tests. Appendix Table 3.E.1 confirms that our results are qualitatively unchanged when controlling for socioeconomic conditions. We measure the

[11] F-tests on the instrument, shown in the table, all exceed the rule of thumb value of 10 for strong instruments.

Table 3.2: IV Tobit With Left-Censoring Estimation Results

	First Stage				
	Dependent variable: Ethnic Diversity				
	(i)	(ii)	(iii)	(iv)	(v)
Kernel Bandwidth:	500 meters	1 km	1.5 km	2 km	3 km
Church Diversity	0.34***	0.33***	0.35***	0.40***	0.56***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]
ln(Population)	-0.06***	-0.05***	-0.05***	-0.05***	-0.05***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	[0.02]	[0.02]	[0.01]	[0.01]	[0.01]
ln(Population)^2	0.01***	0.01***	0.01***	0.01***	0.01***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
	[0.003]	[0.003]	[0.003]	[0.003]	[0.002]
N	6,109	7,931	8,196	8,255	8,258
R^2	0.09	0.12	0.12	0.13	0.17
F-Statistic on Instrument	576.30	992.93	1118.29	1162.90	1666.82
	Second Stage				
	Dependent variable: Number of Victims				
	(i)	(ii)	(iii)	(iv)	(v)
Kernel Bandwidth:	500 meters	1 km	1.5 km	2 km	3 km
Ethnic Diversity	8.71***	7.96***	5.34***	7.36***	4.62***
	(1.05)	(0.96)	(0.98)	(0.96)	(0.79)
ln(Population)	-2.36***	-2.59***	-2.76***	-2.64***	-2.76***
	(0.34)	(0.29)	(0.28)	(0.28)	(0.27)
ln(Population)^2	0.47***	0.51***	0.54***	0.54***	0.54***
	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)
N	6,109	7,931	8,196	8,255	8,258
Wald χ^2	291.65	350.05	341.09	350.24	342.11

**$p<0.05$; ***$p<0.01$

Inference uses spatial autocorrelation-adjusted standard errors when available.

2SLS standard errors reported in (); spatial autocorrelation-adjusted standard errors in []

Regressions include a constant, not shown.

relative affluence of a grid cell using 1971 census data on the quality of housing stock, specifically whether a home had an indoor shower and/or bath; hot water; and toilets. Although violence is higher in less affluent areas, adding any of these controls does not substantially change our estimates of diversity's effect.

Table 3.F.1 considers whether diversity's effect is specific to civilian victimization, as distinct from combat encounters (Balcells 2011; Kalyvas 2006). Segmenting our data according to victim type (civilian or combatant) and comparing the results, the coefficients on diversity are positive and not statistically distinguishable. Diversity's positive effect on conflict intensity is not dependent on victim type, indicating that civilian victimization is not exclusively driving our results.

Table 3.G.1 considers whether ethnic diversity differently affects paramilitaries' and state forces' behavior. We segment our data according to perpetrator type (loyalist, Republican, or British security). All three combatant types killed more victims in areas of greater ethnic diversity.

Table 3.H.1 explores whether ethnically homogenous "strongholds" confound the estimated effect of diversity on violence. It is possible that diverse areas situated next to strongholds are violent because of tactical opportunities – for instance, the ability of militants to find shelter after staging attacks. We use zero-diversity cells as a proxy for strongholds. The estimated effect of diversity is qualitatively unchanged when controlling for each cells log distance to the nearest zero-diversity cell.

Tables 3.I.1-A8 show that our two-stage estimation results are robust to using Poisson instead of a Tobit specification, and that diversity increases violence on both intensive and extensive margins. Our results are also robust to excluding observations nearest churches (Table A9), excluding deaths after 1990 or after 1980 (Table 3.M.1), using raw census data to construct the diversity variable instead of interpolating for missing values (Table 3.N.1), allowing kernel bandwidths to vary for Catholics and Protestants (Table 3.O.1), and aggregating the unit of analysis from 100 meters to 500 meters (Table 3.P.1).

Finally, Table 3.Q.1 separates the diversity variable and instrumented diversity variable into quintiles and allows for nonlinear effects on conflict. Violence increases with diversity throughout the distribution and peaks in the most diverse quintile, a result that holds for both civilian and combatant victims. Although our research design does not determine the precise causal mechanism driving these results, our analysis suggests that the mechanism is not strictly informational. Informational theories (Kalyvas 2006, e.g.) suggest that civilian victimization will peak in contested areas; combatant deaths will peak where one side exerts just enough control to identify the other side's agents (likely in areas with some, but not most, ethnic diversity). Our findings support the first prediction but appear to contradict the second. Informational theories developed in non-ethnic and "conventional" civil wars may not extrapolate directly to ethnic or irregular conflicts. Combatant deaths, at least, may be driven by additional mechanisms: security dilemmas (Posen 1993), hatreds (Kaufmann 1996), homeland mentalities (Toft 2003), or strategic provocations that multiply or intensify combat engagements (Wood 2010).

3.8 Conclusion

Theories of ethnic conflict suggest that in the absence of interethnic social contact or shared economic and social institutions, diverse areas will experience more intense violence. Possible causal mechanisms included heightened grievances, emotional tensions, security dilemmas, and the difficulty of discerning combatants from civilians. Using data from the Troubles in Northern Ireland, we find a strong positive association between local ethnic diversity and conflict intensity, on both intensive and extensive margins. This result substantiates a fundamental assumption underlying social capital and social contact theories of conflict reduction.

Moreover, our paper emphasizes the importance of accounting for local context in modeling conflict. Without first knowing the nature of inter-ethnic social and economic relationships, we cannot know whether to expect more or less violence in diverse areas. This is an important consideration in post-conflict peacebuilding. A debate exists between advocates of ethnic "partition" to prevent further conflict (Kaufmann 1996) and advocates of "ethnic integration" via democratization and power sharing (Hartzell and Hoddie 2003, Sambanis 2000: 482). We do not directly address the question of stability – whether diversity makes future conflict more or less likely. However, our results suggest that in the absence of meaningful social or economic links, persistent diversity will make any future violence more intense. Power sharing and democratization advocates should also encourage contact and the development of interethnic social networks and institutions in labor settings, schools, voluntary associations, and governance bodies. Our analysis supports what community leaders in Northern Ireland have been recommending, based on local wisdom, for several decades (Hughes et al. 2008; Richardson and Gallagher 2011; Tam et al. 2009).

We also make methodological contributions to the study of ethnic politics and violence. We demonstrate the importance of considering endogenous population sorting and measurement error in data from conflict areas. Failing to account for these factors endogenizes measures of diversity to the conflict, biasing results. Our IV estimates reveal that this bias is non-trivial: unadjusted estimates understate diversity's effect size by half. Further, we demonstrate the utility of historical geospatial data to measure population settlement patterns.

Our findings also suggest opportunities for future research. First, our approach – using religious structures as instruments – can alleviate data limitations in the analysis of many ethnic conflicts. Two thirds of countries record no ethnicity data in censuses. Physical structures can proxy for ethnic settlement patterns in these countries (e.g. Turkey, France) and in countries where censuses undercount particular groups (Aktürk 2012; Kopstein and Wittenberg 2011). These physical structures also serve as neighborhood amenities that may reduce ethnic sorting and increase the risk of conflict via their effect on local demographics (Field et al. 2008). Second, our research design might be extended to qualitatively different contexts, such as ideological wars or ethnic civil wars with conventional front lines. In conventional, ideological civil wars, belligerents strategically eliminate the opposing side's civilian supporters from territory they conquer (Balcells 2011). Using physical structures to instrument for pre-war diversity would enable research on behind-the-lines civilian victimization in conventional ethnic civil wars. It could also enable research on irregular ideological civil wars, if appropriate physical instruments exist for ideological diversity.

Chapter 3, in part is currently being prepared for submission for publication of the material. McCord, Gordon, Joseph Brown, and Paul Zachary.

Bibliography

Acharya, Avidit, Matthew Blackwell, and Maya Sen. 2016. "The political legacy of American slavery." *Journal of Politics* 78 (3): 621–641.

Adida, Claire, David Laitin, and Marie-Anne Valfort. 2015. ""One Muslim is Enough!" Evidence from a Field Experiment in France." *Annals of Economics and Statistics* .

Aktürk, Şener. 2012. *Regimes of ethnicity and nationhood in Germany, Russia, and Turkey*. New York, NY: Cambridge University Press.

Allport, Gordon W. 1954. *The Nature of Prejudice*. Reading, MA: Addison-Wesley Publishing.

Balcells, Laia. 2011. "Continuation of Politics by Two Means: Direct and Indirect Violence in Civil War." *Journal of Conflict Resolution* 55 (3): 397–422.

Balcells, Laia, Lesley-Ann Daniels, and Abel Escribà-Folch. 2016. "The determinants of low-intensity intergroup violence: The case of Northern Ireland." *Journal of Peace Research* 53 (1): 33–48.

Bardon, Jonathan. 1982. *Belfast: An Illustrated History*. Dundonald, Northern Ireland: The Blackstaff Press.

Barritt, Denis P., and Charles F. Carter. 1962. *The Northern Ireland Problem: A Study In Group Relations*. London: Oxford University Press.

Bell, J. Bowyer. 1997. *The Secret Army: The IRA*. New Brunswick, New Jersey: Transaction Publishers.

Bhavnani, Ravi, Karsten Donnay, Dan Miodownik, Maayan Mor, and Dirk Helbing. 2014. "Group segregation and urban violence." *American Journal of Political Science* 58 (1): 226–245.

Blalock, H. M. 1967. *Toward a theory of minority-group relations*. New York, NY: Capricorn Books.

Boal, Frederick W. 1982. "Segregation and Mixing: Space and Residence in Belfast." In *Integration and Division: Geographical Perspectives on the Northern Ireland Problem*, ed. Frederick W. Boal, and J. Neville H. Douglas. New York, NY: Academic Press.

Boal, Fredrick W. 1969. "Territoriality on the shankill-falls divide, Belfast." *Irish Geography* 6 (1): 30–50.

Brass, Paul R. 1997. *Theft of an idol: Text and context in the representation of collective violence*. Princeton, NJ: Princeton University Press.

Brewer, John, Higgins Gareth, and Francis Teeney. 2011. *Religion, Civil Society, and Peace in Northern Ireland*. London: Oxford University Press.

CAIN. 2014. *Conflict and Politics in Northern Ireland*. Derry/Londonderry, Northern Ireland: University of Ulster.

Cederman, Lars-Erik, and Kristian Skrede Gleditsch. 2009. “Introduction to Special Issue on “Disaggregating Civil War”.” *Journal of Conflict Resolution* 53 (4): 487–495.

Cederman, Lars-Erik, Nils B Weidmann, and Kristian Skrede Gleditsch. 2011. “Horizontal inequalities and ethnonationalist civil war: A global comparison.” *American Political Science Review* 105 (03): 478–495.

Chandra, Kanchan. 2012. *Constructivist Theories of Ethnic Politics*. New York, NY: Oxford University Press chapter Introduction.

Clack, B., J. Dixon, and C. Tredoux. 2005. “Eating together apart: Patterns of segregation in a multi-ethnic cafeteria.” *Journal of Community and Applied Social Psychology* 15: 1–16.

Collier, Paul, and Anke Hoeffler. 2004. “Greed and grievance in civil war.” *Oxford Economic Papers* 56 (4): 563–595.

Conley, Timothy G. 1999. “GMM estimation with cross sectional dependence.” *Journal of Econometrics* 92 (1): 1–45.

Coogan, Tim Pat. 1993. *The IRA*. London: Harper Collins.

Cunningham, Kathleen G., and Nils B. Weidmann. 2010. “Shared Space: Ethnic Groups, State Accommodation, and Localized Conflict.” *International Studies Quarterly* 54 (4): 1035–1054.

Dancygier, Rafaela M. 2010. *Immigration and conflict in Europe*. New York, NY: Cambridge University Press.

Darby, John. 1978. “Northern Ireland: Bonds and Breaks in Education.” *British Journal of Education Studies* 26 (3): 215–223.

Dillon, Martin. 1999. *The Shankill Butchers: The Real Story of Cold-Blooded Mass Murder*. New York, NY: Routledge.

Dixon, J., C. Tredoux, K. Durrheim, G. Finchilescu, and B. Clack. 2008. “The inner citadels of the color line: Mapping the micro-ecology of racial segregation in everyday life spaces.” *Social and Personality Psychology Compass* 2: 1547–1569.

Ellingsen, Tanja. 2000. “Colorful community or ethnic witches’ brew? Multiethnicity and domestic conflict during and after the cold war.” *Journal of Conflict Resolution* 44 (2): 228–249.

Fearon, James D, and David D Laitin. 2003. “Ethnicity, insurgency, and civil war.” *American Political Science Review* 97 (01): 75–90.

Fermanagh Herald. 1971. “Census forms burned in Belfast streets.”.

Field, Erica, Matthew Levinson, Rohini Pande, and Sujata Visaria. 2008. "Segregation, rent control, and riots: The economics of religious conflict in an Indian city." *American Economic Review* 98 (2): 505–510.

Gallagher, James. 1971. "Police will not be told." *Irish Independent* p. 15.

Glaeser, Edward L. 2005. "The political economy of hatred." *Quarterly Journal of Economics* 120 (1): 45–86.

Griffiths, Hywel. 1971. *FLIGHT: A Report on Population Movement in Belfast during August, 1971*. Belfast: Northern Ireland Community Relations Commission.

Hancock, Landon. 1998. *Northern Ireland: Troubles Brewing*. Derry, UK: CAIN, University of Ulster.

Hardin, Russell. 1995. *One for All: The Logic of Group Conflict*. Princeton, New Jersey: Princeton University Press.

Hartzell, Caroline, and Matthew Hoddie. 2003. "Institutionalizing Peace: Power Sharing and Post-Civil War Conflict Management." *American Journal of Political Science* 47 (2): 318–332.

Hechter, Michael. 1977. *Internal colonialism: The Celtic fringe in British national development, 1536-1966*. Vol. 197 Berkeley, CA: University of California Press.

Hepburn, A.C. 1996. *A Past Apart: Studies in the History of Catholic Belfast 1850-1950*. Belfast: Ulster Historical Foundation.

Hewstone, Miles, and Hermann Swart. 2011. "Fifty-odd years of inter-group contact: From hypothesis to integrated theory." *British Journal of Social Psychology* 50: 374–386.

Hughes, Joanne, Andrea Campbell, Miles Hewstone, and Ed Cairns. 2008. "What's there to fear? A comparative study of responses to the out-group in mixed and segregated areas of Belfast." *Peace & Change* 33 (4): 522–548.

Hughes, Kathleen. 1966. *The Church In Early Irish Society*. Ithaca, NY: Cornell University Press.

Irvine, Maurice. 1991. *Northern Ireland: Faith and Faction*. New York, NY: Routledge.

Jha, Saumitra. 2013. "Trade, institutions, and ethnic tolerance: Evidence from South Asia." *American Political Science Review* 107 (04): 806–832.

Kalyvas, Stathis. 2006. *The Logic of Violence in Civil War*. New York, NY: Cambridge University Press.

Kalyvas, Stathis N. 2005. "Warfare In Civil Wars." In *Rethinking the Nature of War*, ed. Isabelle Duyvesteyn, and Jan Angstrom. Abington, UK: Frank Cass.

Kasara, Kimuli. 2016. "Does Local Ethnic Segregation Lead to Violence?: Evidence from Kenya." *Quarterly Journal of Political Science* (forthcoming).

Kaufmann, Chaim D. 1996. "Possible and Impossible Solutions to Ethnic Wars." *International Security* 20 (4): 136–175.

Klašnja, Marko, and Natalija Novta. 2014. "Segregation, Polarization, and Ethnic Conflict." *Journal of Conflict Resolution* .

Knox, H.M. 1973. "Religious Segregation in the Schools of Northern Ireland." *British Journal of Education Studies* 21 (3): 307–312.

Kopstein, Jeffrey S, and Jason Wittenberg. 2011. "Deadly communities: Local political milieus and the persecution of Jews in occupied Poland." *Comparative Political Studies* 44 (3): 259–283.

Lemmer, Gunnar, and Ulrich Wagner. 2015. "Can we really reduce ethnic prejudice outside the lab? A meta-analysis of direct and indirect contact interventions." *European Journal of Social Psychology* 45: 151–168.

Loyle, Cyanne E, Christopher Sullivan, and Christian Davenport. 2014. "The Northern Ireland Research Initiative: Data on the Troubles from 1968 to 1998." *Conflict Management and Peace Science* 31 (1): 94–106.

Lynch, John. 1998. *A Tale of Three Cities: Comparative Studies in Working-Class Life*. London: Macmillan Press.

McGee, Caroline M. 2013. "'A noble Church in the most Catholic quarter of a bitterly Protestant and Presbyterian city: The Church of the Most Holy Redeemer, Clonard, West Belfast'." In *Belfast: The Emerging City, 1850-1914*, ed. Olwen Purdue. Dublin: Irish Academic Press.

McKeown, Ciaran. 1971. "North census row grows." *Irish Press* p. 1.

McKittrick, David, Seamus Kelters, Brian Feeney, Chris Thornton, and David McVea. 2001. *Lost Lives: The stories of the men, women and children who died as a result of the Northern Ireland troubles*. Random House.

Moloney, Ed. 2002. *A Secret History of the IRA*. New York, NY: W.W. Norton and Company.

Moloney, Ed. 2010. *Voices from the Grave: Two Men's War in Ireland*. New York, NY: Public Affairs.

Moxon-Browne, Edward. 1991. *Social Attitudes in Northern Ireland: The First Report*. Belfast: Blackstaff Press chapter National Identity in Northern Ireland.

Murray, Dominic. 1985. *Worlds Apart: Segregated Schools In Northern Ireland*. Belfast: Appletree Press.

Northern Ireland Statistics and Research Agency. 2005. "Urban-Rural Classification (2005)." Online.

O'Connor, Tom. 1971. "Paisley urges census action." *Irish Press* p. 3.

O'Hearn, Denis. 1983. "Catholic Grievances, Catholic Nationalism: A Comment." *The British Journal of Sociology* 34 (3): 438–445.

Petersen, Roger D. 2002. *Understanding ethnic violence: Fear, hatred, and resentment in twentieth-century Eastern Europe*. New York, NY: Cambridge University Press.

Pettigrew, Thomas F, and Linda R Tropp. 2006. "A meta-analytic test of intergroup contact theory." *Journal of personality and social psychology* 90 (5): 751.

Pettigrew, Thomas F, and Linda R Tropp. 2008. "How does intergroup contact reduce prejudice? Meta-analytic tests of three mediators." *European Journal of Social Psychology* 38 (6): 922–934.

Posen, Barry R. 1993. "The Security Dilemma and Ethnic Conflict." *Survival* 35 (1): 27–47.

Putnam, Robert D. 2000. *Bowling Alone: The Collapse and Revival of American Community*. New York, NY: Simon and Schuster.

Richardson, Norman, and Tony Gallagher. 2011. *Education for Diversity and Mutual Understanding: The Case of Northern Ireland*. New York: Peter Lang.

Robinson, Gillian. 1992. *Cross-Community Marriage in Northern Ireland*. Belfast: Centre for Social Research.

Rowthorn, Bob. 1981. "Northern Ireland: an economy in crisis." *Cambridge Journal of Economics* 5 (1): 1–31.

Sambanis, Nicholas. 2000. "Partition As A Solution to Ethnic War: An Empirical Critique of the Theoretical Literature." *World Politics* 52 (4): 437–483.

Shuttleworth, Ian, and Christopher Lloyd. 2007. "Linking Northern Ireland Census of Population Data, 1971-2001." *UK Data Service* .

Staiger, Douglas, and James H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65 (3): 557–586.

Sunday Independent. 1971. "Bombs blast in Belfast.".

Tam, Tania, Miles Hewstone, Jared Kenworthy, and Ed Cairns. 2009. "Intergroup trust in Northern Ireland." *Personality and Social Psychology Bulletin* 35 (1): 45–59.

Toft, Monica Duffy. 2003. *The Geography of Ethnic Violence*. Princeton, NJ: Princeton University Press.

UVF. 1971. "Ulster Volunteer Force Recruiting Circular (1971)." In *Bigotry and Blood: Documents on the Ulster Troubles*, ed. Charles Carlton. Chicago, IL: Nelson Hall.

Varshney, Ashutosh. 2001. "Ethnic Conflict and Civil Society: India and Beyond." *World Politics* 53 (3): 362–398.

Weidmann, Nils B, and Idean Salehyan. 2013. "Violence and ethnic segregation: a computational model applied to Baghdad." *International Studies Quarterly* 57 (1): 52–64.

Wilkinson, Steven I. 2006. *Votes and violence: Electoral competition and ethnic riots in India*. New York, NY: Cambridge University Press.

Wood, Reed M. 2010. "Rebel Capability and Strategic Violence Against Civilians." *Journal of Peace Research* 47 (5): 601–614.

Appendix

3.A Summary Statistics

Table 3.A.1: Summary Statistics

	Obs	Mean	Std. Dev.	Min	Max
Victims	8,258	0.20	0.90	0	23
Ethnic Diversity	5,451	0.27	0.27	0	1
Ethnic Diversity (Interpolated)	8,258	0.28	0.25	0	1
Population	8,258	71.19	76.66	1	1,438
Church Diversity Instrument (3 km Kernel)	8,258	0.30	0.19	0	1
Distance to Nearest Protestant Church (m)	8,258	463.01	325.46	10	2,075
Distance to Nearest Catholic Church (m)	8,258	1110.81	877.17	11	5,846
Population without Hot Water (%)	5,451	0.09	0.19	0	.98
Population without Bath or Shower in Home (%)	5,451	0.14	0.29	0	1
Population without Inside Toilet (%)	5,451	0.14	0.29	0	1

Table 3.A.1 presents summary statistics on the variables used in the analysis. There are 8,258 locations within Belfast included in the 1971 census. The data on victims have been mapped to those locations. Only 5,451 locations have data on religious composition and on assets. We interpolate the religious composition in order to use all 8,258 locations. The distance to church variables and the population asset variables are employed in the robustness estimations. Distance to nearest church is calculated with GIS. The asset variables are taken from the 1971 census.

3.B Map of 1971 Census Data in Belfast

As Figure A1 shows, inhabited grid cells from the 1971 census align quite closely with Belfast's contemporary boundaries.

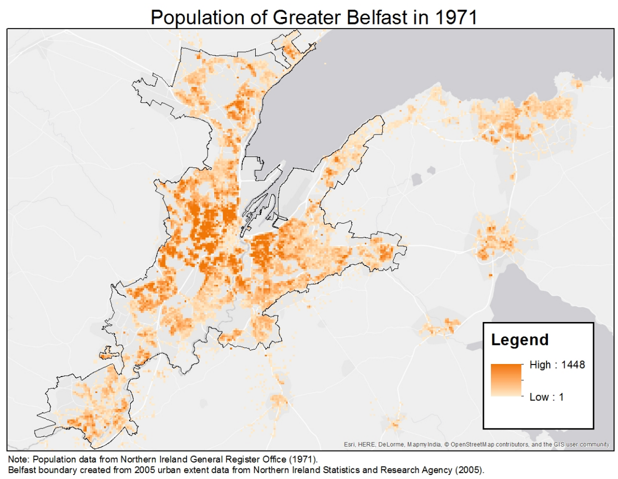

Figure A1: Population of Greater Belfast in 1971. Darker areas are more densely populated, while gray areas have no population. The dark line is the urban boundary of Belfast in 2005.

3.C Archival map vs. Interpolation

Figure A1 shows the result of the interpolation for the highly contested area around Divis Street.

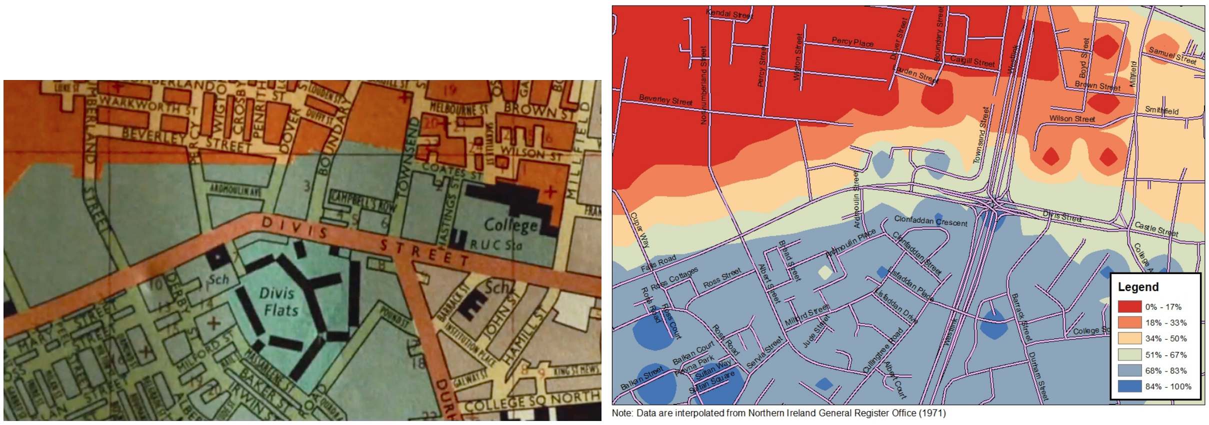

Figure A1: **Left:** Archival map from 1969, prior to ethnic riots. It shows the religion of inhabitants along Divis Street. Red areas are majority Protestant and blue are majority Catholic. **Right:** The same area (with current streets overlaid) with our interpolated measure of religious demographics based on the 1971 census.

3.D Growth of Belfast

Figure A1 shows Belfast's rapid growth during the industrialization of the 19th and 20th Centuries.

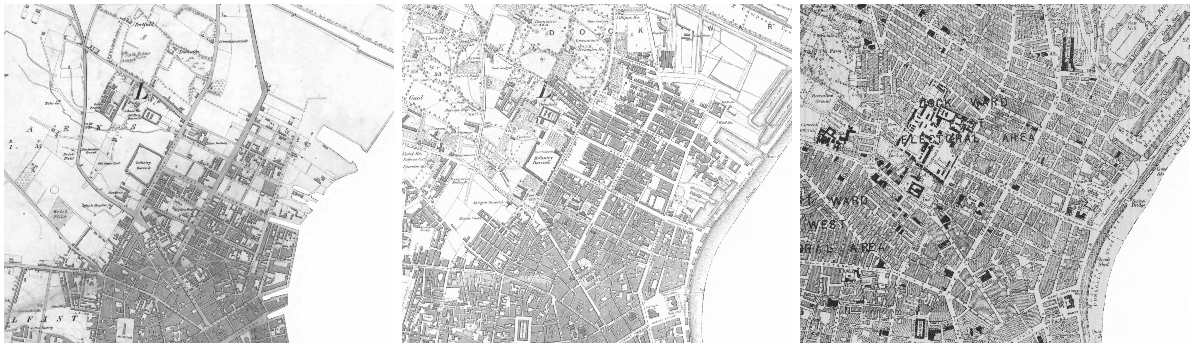

Figure A1: **Left:** The Dock Ward area of Belfast in 1831. In this period, the northernmost extent of urban area was Great Georges Street. **Center:** The same ward in 1858. **Right:** The same area in 1931. Images Crown copyright 2015, used with permission from Land & Property Services Northern Ireland.

3.E Controlling for Socioeconomic Status

Socioeconomic conditions might confound our results if wealth were correlated with violence and church diversity, and if wealth were driving violence levels. In that case, our specifications would be incorrectly assigning wealth's effect to ethnic diversity. To account for this possibility, we use 1971 census data on the quality of housing stock in each grid cell. We focus on housing because the 1971 census did not record respondents' income, forcing us to rely on a proxy. We use measures of whether homes had indoor 1) hot water; 2) shower and/or bath; and 3) toilets. Inhabitants in areas without these facilities are likely poorer than those who have these plumbing amenities. At first glance, the correlation between socioeconomic conditions and ethnic diversity is low. For instance, the correlation between ethnic diversity and the proportion of people without a shower and/or bath is -0.02. Table 3.E.1 shows that violence is positively associated with a lack of household assets. Nevertheless, adding any of these controls does not substantially change our estimates of diversity's effect.

3.F Segmenting Data by Victim Type

We explore the importance of civilian victimization by looking for differences in the association between ethnic diversity and civilian vs. non-civilian deaths. The table below shows the second-stage regression results separately estimated for the two victim types. The two coefficients on diversity are not statistically different from one another. This suggests that the dynamics linking local ethnic diversity to deaths during the Troubles were not specific to civilian victimization.

Table 3.E.1: IV Tobit with Socioeconomic Conditions

	First Stage		
	Dependent variable: Ethnic Diversity		
	(i)	(ii)	(iii)
Church Diversity	0.32^{***}	0.32^{***}	0.32^{***}
	(0.01)	(0.01)	(0.01)
	[0.03]	[0.03]	[0.03]
No Hot Water	-0.09^{***}		
	(0.02)		
	[0.05]		
No Bath/Shower		-0.08^{***}	
		(0.01)	
		[0.03]	
No Inside Toilet			-0.10^{***}
			(0.01)
			[0.03]
N	5,386	5,386	5,386
R^2	0.11	0.12	0.12
F-Statistic on Instrument	552.11	558.61	563.17
	Second Stage		
	Dependent variable: Number of Victims		
	(i)	(ii)	(iii)
Ethnic Diversity	6.58^{***}	6.59^{***}	6.67^{***}
	(1.09)	(1.08)	(1.08)
No Hot Water	4.80^{***}		
	(0.43)		
No Bath/Shower		3.64^{***}	
		(0.31)	
No Inside Toilet			3.91^{***}
			(0.32)
N	5,386	5,386	5,386
Wald χ^2	359.35	369.61	370.01

$^{**}p<0.05$; $^{***}p<0.01$

All regressions use a 1km kernel bandwidth for church diversity calculation.

2SLS standard errors reported in () and spatial autocorrelation

-adjusted standard errors in []

Regressions include constant, ln(Population), and $\ln(\text{Population})^2$, not shown.

Table 3.F.1: Robustness of IV Tobit Estimates to Subsetting on Victim Type

	Second Stage	
	Dependent variable: Number of Victims	
	(i)	(ii)
Victim Type:	Civilians	Non-Civilians
Ethnic Diversity	6.14^{***}	6.79^{***}
	(0.89)	(1.02)
ln(Population)	-2.20^{***}	-1.81^{***}
	(0.27)	(0.31)
ln(Population)^2	0.42^{***}	0.36^{***}
	(0.04)	(0.05)
N	7,931	7,931
Wald chi^2	244.24	304.17

**p<0.05; ***p<0.01

2SLS standard errors reported in ()

Note: First stage results the same as in Table 3.2.

All regressions use a 1km kernel bandwidth for church diversity calculation.

Regressions include a constant, not shown.

3.G Perpetrator Type

We also assess the robustness of our findings when segmenting the victim data by perpetrator type. Table 3.G.1 shows the second-stage results of three Tobit regressions estimated separately on deaths attributed to Republican paramilitaries, loyalist paramilitaries, and British security forces. In all three cases, the relationship between local diversity and number of deaths is positive and strongly significant. The regressions estimated on deaths attributed to Republicans and British security forces yield similar coefficients on the instrumented diversity variable. The regression estimated on deaths attributed to loyalists yields a coefficient of lesser magnitude, smaller by roughly two-thirds. All three perpetrator types killed more victims in diverse areas. The less pronounced association between diversity and violence, looking at loyalist killings only, may stem from the actions of certain paramilitary and vigilante groups (the Shankill Butchers, for example) who carried out sectarian reprisal killings in majority Catholic neighborhoods (Dillon 1999).

Table 3.G.1: Robustness of IV Tobit Estimates to Subsetting on Perpetrator Type

	Second Stage		
	Dependent variable: Number of Victims		
	(i)	(ii)	(iii)
Perpetrator Type:	Republican	Loyalist	British Security
Ethnic Diversity	9.01***	3.31***	8.09***
	(1.11)	(0.98)	(1.56)
ln(Population)	-1.98***	-2.01***	-1.34**
	(0.33)	(0.29)	(0.54)
ln(Population)^2	0.39***	0.38***	0.34***
	(0.05)	(0.04)	(0.07)
N	7,931	7,931	7,931
Wald *chi*^{2}	182.31	158.38	86.48

**p<0.05; ***p<0.01

2SLS standard errors reported in ()

Note: First stage results the same as in Table 3.2.

All regressions use a 1km kernel bandwidth for church diversity calculation.

Regressions include a constant, not shown.

3.H Distance from Stronghold

Ethnic “strongholds,” or neighborhoods under the exclusive control of one ethnic group, may present tactical opportunities for combatants, who can quickly retreat into their respective strongholds after engaging in violence. Strongholds correlate with diversity, so any effect of strongholds on violence could confound our estimate of diversity’s effect. To assess this possibility, we use ethnically homogenous cells (those with a zero diversity score) to proxy for the location of strongholds. We then control for each grid cell’s log distance to its nearest stronghold.

The coefficient on the distance-to-stronghold variable is -0.59, statistically significant at the 99% level. Log distance from the nearest homogeneous cell is inversely associated with the number of victims, suggesting that strongholds do intensify violence in nearby grid cells. More important to us, however, the coefficient on diversity remains positive and statistically significant at the 99% level. In fact, the coefficient on diversity increases to 10.91, versus 7.96 in the basic model with the equivalent 1 kilometer bandwidth (see Table 2). As in all model specifications we examine, diversity exerts a strong intensifying effect on violence, regardless of the controls used.

Table 3.H.1: Robustness of IV Tobit Estimates to Distance from Stronghold

	First Stage *Dependent variable: Diversity*
	(i)
Kernel Bandwidth:	1 km
Church Diversity	0.22***
	(0.01)
ln(Population)	0.04***
	(0.01)
ln(Population)^2	-0.00***
	(0.00)
ln(Distance to Stronghold)	0.07***
	(0.001)
N	7,931
R^2	0.39
F-Statistic on Instrument	606.45
	Second Stage *Dependent variable: Number of Victims*
	(i)
Ethnic Diversity	10.91***
	(1.50)
ln(Population)	-3.22***
	(0.30)
ln(Population)^2	0.58***
	(0.04)
ln(Distance to Stronghold)	-0.59***
	(0.12)
N	7,931
Wald χ^2	349.31

**p<0.05; ***p<0.01

Regressions include a constant, not shown.

3.I IV Poisson

As an alternative to the Tobit specification in the second stage, Table 3.I.1 uses the same first stage estimation, instrumenting ethnic diversity with church diversity, with a Poisson specification in the second stage. (This specification is sometimes more appropriate for count data, but it makes strong assumptions about the distribution of the data.) The second stage coefficient on ethnic diversity is significant at all bandwidths and is consistently larger than the corresponding coefficient of 1.6 in the Poisson regression of Table 3.1. These coefficients suggest that maximally diverse cells will have roughly 2-5 more deaths than cells which are completely homogeneous.

Table 3.I.1: IV Poisson Estimation Results

	Second Stage				
	Dependent variable: Number of Victims				
	(i)	(ii)	(iii)	(iv)	(v)
Kernel Bandwidth:	500 meters	1 km	1.5 km	2 km	3 km
Ethnic Diversity	4.69^{***}	5.47^{***}	2.78^{***}	3.54^{***}	1.98^{***}
	(1.29)	(2.04)	(0.49)	(0.56)	(0.41)
ln(Population)	-0.79^{***}	-0.91^{***}	-1.01^{***}	-0.97^{***}	-1.03^{***}
	(0.18)	(0.17)	(0.15)	(0.15)	(0.14)
$\ln(\text{Population})^2$	0.16^{***}	0.18^{***}	0.20^{***}	0.19^{***}	0.20^{***}
	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
N	6,109	7,931	8,196	8,255	8,258

$^{**}p<0.05$; $^{***}p<0.01$

Note: First stage results the same as in Table 3.2.

Regressions include a constant, not shown.

3.J Intensive Margin Results

Here, as in Table 3.1, we use the IV approach to estimate the effect of ethnic diversity on the intensive and extensive margins of violence. Our results on the intensive margin (limiting the sample to locations with non-zero deaths) are presented in Table 3.J.1 for the various kernel bandwidths of church density. The first stage shows that the instrument is positively correlated with ethnic diversity across all kernel bandwidths, and is significant at the 99% confidence level using standard or spatially-adjusted standard errors. F-statistics are above 10 in all regressions, indicating that the instrument is strong. The second stage coefficients are significant at all bandwidths except 3 km (not surprising given that the first stage is weaker at this bandwidth). The coefficients are consistent in magnitude, ranging from 1.5 to 2.8. This suggests that a maximally diverse cell will experience 2-3 more deaths on average than homogenous cells, conditional on having at least one death. The corresponding coefficient in column (iv) of Table 3.1 is 1.05. These results suggest not only that ethnic diversity plays a role in determining the intensity of conflict, but that unadjusted OLS regressions may underestimate the importance of the relationship.

Table 3.J.1: IV Estimation Results for Victims on the Intensive Margin

	First Stage				
	Dependent variable: Ethnic Diversity				
	(i)	(ii)	(iii)	(iv)	(v)
Kernel Bandwidth:	500 meters	1 km	1.5 km	2 km	3 km
Church Diversity	0.27***	0.33***	0.33***	0.27***	0.22***
	(0.04)	(0.04)	(0.05)	(0.05)	(0.06)
	[0.05]	[0.05]	[0.07]	[0.06]	[0.07]
ln(Population)	-0.03	-0.02	-0.04	-0.05	-0.05
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
	[0.03]	[0.03]	[0.03]	[0.03]	[0.03]
ln(Population)^2	0.0001	-0.001	0.002	0.003	0.004
	(0.01)	(0.004)	(0.004)	(0.005)	(0.005)
	[0.01]	[0.005]	[0.005]	[0.005]	[0.01]
N	709	778	784	785	785
R^2	0.08	0.11	0.07	0.05	0.03
F-Statistic on Instrument	50.39	79.68	47.91	25.89	14.30
	Second Stage				
	Dependent variable: Number of Victims				
Kernel Bandwidth:	500 meters	1 km	1.5 km	2 km	3 km
Ethnic Diversity	2.74	2.84***	2.38***	2.81**	1.49
	(1.13)	(0.91)	(1.13)	(1.53)	(2.00)
	[1.48]	[1.00]	[0.92]	[1.24]	[1.53]
ln(Population)	-0.30	-0.38	-0.40	-0.38	-0.44
	(0.31)	(0.27)	(0.25)	(0.26)	(0.26)
	[0.30]	[0.24]	[0.23]	[0.23]	[0.23]
ln(Population)^2	0.05	0.07**	0.07**	0.07**	0.07**
	(0.04)	(0.04)	(0.03)	(0.04)	(0.03)
	[0.04]	[0.03]	[0.03]	[0.03]	[0.03]
N	709	778	784	785	785

**p<0.05; ***p<0.01

Inference uses spatial autocorrelation-adjusted standard errors.

2SLS errors reported in () and spatial autocorrelation-adjusted standard errors in []

Regressions include a constant, not shown.

3.K Extensive Margin Results

The results on the extensive margin are presented in Table 3.K.1 for the various kernel bandwidths of church density. The second stage coefficients are significant at all bandwidths. The magnitude of the coefficient is consistent across kernel bandwidths, ranging from 1.07 to 1.98. (For comparison, a probit regression ignoring the endogeneity of the diversity variable yields a coefficient of 1.46.)

Table 3.K.1: IV Probit Results for Extensive Margin of Violence Location

	Second Stage				
	Dependent variable: I(Deaths > 0)				
Kernel Bandwidth:	500 meters	1 km	1.5 km	2 km	3 km
Ethnic Diversity	1.98^{***}	1.76^{***}	1.20^{***}	1.66^{***}	1.07^{***}
	(0.20)	(0.20)	(0.22)	(0.20)	(0.18)
ln(Population)	-0.57^{***}	-0.62^{***}	-0.68^{***}	-0.63^{***}	-0.68^{***}
	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)
$\ln(\text{Population})^2$	0.12^{***}	0.12^{***}	0.13^{***}	0.13^{***}	0.13^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
N	6,109	7,931	8,196	8,255	8,258

$^{**}p<0.05$; $^{***}p<0.01$
Note: First stage results the same as in Table 3.2.
Regressions include constant, not shown.

3.L Exclusion Restriction

Our use of Catholic and Protestant church densities to construct an instrumental variable for diversity would violate the exclusion restriction if churches relate to violence through any channel other than neighborhood demographics. We argue in the text that churches were unlikely to cause an increase in violence, because Catholic and Protestant clergy denounced violence consistently throughout the conflict. Paramilitaries ignored these appeals, however, and violent confrontations occasionally spilled over onto church grounds. We conclude that the presence of churches neither encouraged nor discouraged violence.

For thoroughness, however, we empirically assess how sensitive our results are to excluding data in the immediate vicinity of churches. We re-estimate the main model using a 1 kilometer church density kernel – column (ii) in Table 3.2 – while dropping data points within 100 meters, 250 meters and 500 meters of churches. Our results are robust to excluding these observations.

These findings are presented in Table 3.L.1. The first-stage coefficients on the instrument remain consistent with the corresponding 0.33 estimate from Table 3.2 and are significant at the 99% level. The instrument remains strong, as indicated by F-statistics greater than 10. The second stage coefficients on diversity remain statistically significant, and their magnitudes are even larger than the second stage coefficient in Table 3.2. We maintain the more conservative results in Table 3.2 as our preferred specification.

Table 3.L.1: IV Tobit Estimation Dropping Observations Near Churches

	First Stage		
	Dependent variable: Ethnic Diversity		
	(i)	(ii)	(iii)
Excluded Area Around Church:	100 m	250 m	500 m
Church Diversity	0.31***	0.24***	0.19***
	(0.01)	(0.01)	(0.02)
	[0.03]	[0.03]	[0.04]
ln(Population)	-0.06***	-0.08***	-0.12***
	(0.01)	(0.01)	(0.02)
	[0.02]	[0.02]	[0.02]
ln(Population)^2	0.01***	0.01***	0.02***
	(0.002)	(0.002)	(0.003)
	[0.003]	[0.003]	[0.004]
N	7,294	5,111	2,184
R^2	0.11	0.09	0.09
F-Statistic on Instrument	823.00	390.47	102.63
	Second Stage		
	Dependent variable: Number of Victims		
	(i)	(ii)	(iii)
Excluded Area Around Church:	100 m	250 m	500 m
Ethnic Diversity	8.68***	11.90***	14.51***
	(1.04)	(1.73)	(3.57)
ln(Population)	-2.65***	-2.37***	-1.19
	(0.30)	(0.43)	(0.75)
ln(Population)^2	0.51***	0.45***	0.19
	(0.04)	(0.07)	(0.13)
N	7,294	5,111	2,184
Wald χ^2	332.67	182.36	42.09

**p<0.05; ***p<0.01
2SLS standard errors reported in () and spatial autocorrelation-adjusted standard errors in []
All regressions use a 1km kernel bandwidth for church diversity calculation.
Regressions include a constant, not shown.

3.M Time Period Effects

Deaths attributable to the Troubles span several decades – from the 1960s to the early 2000s. One potential concern regarding our results is that they could be driven by conflict during certain years. For instance, 1969-1979 were the most intense years of the conflict, accounting for roughly 60% of the deaths (2,112 of 3,532). The largest ethnic dislocations from mixed neighborhoods to homogeneous neighborhoods also occurred during the early years of the conflict. Spatial diversity during the late 1960s may be more closely related to conflict in the 1970s than to conflict in the 1980s and 1990s.

A reasonable robustness check on our findings is to determine that they do not depend on deaths during these later decades. We therefore re-estimate our main model (using the 1 kilometer kernel bandwidth) omitting violent deaths in the 1980s and 1990s. These results are presented in Table 3.M.1. Our main findings are robust to excluding deaths that occurred after 1980 (the first column) and deaths that occurred after 1990 (second column). The second stage coefficients on diversity are 7.45 and 8.18, consistent with the 7.96 coefficient in the corresponding Tobit regression in Table 3.2. These results indicate that our findings are not driven by deaths in the 1980s and 1990s, long after the 1971 diversity measure was taken.

Table 3.M.1: Robustness of IV Tobit Estimates to Excluding Later Deaths

	Second Stage	
	Dependent variable: Number of Victims	
	(i)	(ii)
Excluding Deaths After:	1980	1990
Ethnic Diversity	7.45***	8.18***
	(1.06)	(1.00)
ln(Population)	-2.19***	-2.32***
	(0.33)	(0.31)
ln(Population)^2	0.45***	0.47***
	(0.05)	(0.04)
N	7,931	7,931
Wald chi^2	244.24	304.17

**p<0.05; ***p<0.01

2SLS standard errors reported in ()

Note: First stage results the same as in Table 3.2.

All regressions use a 1km kernel bandwidth for church diversity calculation.

Regressions include a constant, not shown.

3.N Non-interpolated data

As explained in the main body of the paper, we use ordinary Kriging to replace missing data from the 1971 census with interpolated values. To test whether the interpolation is driving our results, we re-estimate the main IV Tobit model using only cells that reported religion data in the 1971 census. Table 3.N.1 shows that our results remain robust and consistent with the main IV Tobit results in Table 3.2.

Table 3.N.1: Non-Interpolated IV Tobit With Left-Censoring Estimation Results

	First Stage				
	Dependent variable: Diversity				
	(i)	(ii)	(iii)	(iv)	(v)
Kernel Bandwidth:	500 meters	1 km	1.5 km	2 km	3 km
Church Diversity	0.30***	0.31***	0.35***	0.38***	0.50***
	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)
	[0.03]	[0.03]	[0.03]	[0.03]	[0.04]
ln(Population)	0.42***	0.44***	0.33***	0.30***	0.26***
	(0.07)	(0.07)	(0.06)	(0.06)	(0.06)
	[0.11]	[0.10]	[0.10]	[0.10]	[0.10]
ln(Population)^2	-0.04***	-0.04***	-0.03***	-0.03***	-0.03***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]
N	4,396	5,386	5,451	5,451	5,451
R^2	0.07	0.11	0.12	0.11	0.14
F-Statistic on Instrument	292.66	538.87	624.09	564.40	764.80
	Second Stage				
	Dependent variable: Number of Victims				
	(i)	(ii)	(iii)	(iv)	(v)
Kernel Bandwidth:	500 meters	1 km	1.5 km	2 km	3 km
Ethnic Diversity	8.05***	7.18***	4.57***	6.92***	4.41***
	(1.32)	(1.12)	(1.11)	(1.16)	(0.97)
ln(Population)	-6.61***	-8.05***	-7.10***	-8.20***	-7.12***
	(1.85)	(1.76)	(1.74)	(1.78)	(1.74)
ln(Population)^2	0.95***	1.11***	1.02***	1.13***	1.03***
	(0.20)	(0.19)	(0.19)	(0.19)	(0.19)
N	4,396	5,386	5,451	5,451	5,451
Wald χ^2	245.93	297.33	295.12	295.61	294.64

**p<0.05; ***p<0.01

2SLS standard errors reported in () and spatial autocorrelation-adjusted standard errors in []

Inference uses spatial autocorrelation-adjusted standard errors when available.

Regressions include a constant, not shown.

3.O Varying Kernel Density across Ethnic Groups

A potential concern with the construction of our instrument is that the "amenity distance" of churches may differ for Protestants and Catholics. In that case, our use of identical kernel bandwidths for Protestant and Catholic churches might not be consistent with the actual settlement patterns of Protestants and Catholics. This would weaken our instrument, but not invalidate it. Instrument strength is not a concern in our analysis as evidenced by the first stage F-statistics throughout. As a robustness check, we use the mean distance to churches (463 meters for Protestants; 1,100 meters for Catholics) as a guide to construct a version of the instrument with different kernel sizes for Protestants and Catholics (500m for the former, 1km for the latter). Our results, using this alternative instrument in the IV Tobit regression, are shown in Table 3.O.1.

The first-stage coefficient on the instrument is significant at the 99% level, and the F-statistic indicates that the instrument is strong. The coefficient on diversity in the second stage (10.26) is consistent with the second-stage coefficients in Table 2, and is significant at the 99% level. Thus, allowing different bandwidths across Catholics and Protestants leaves our results essentially unchanged.

A more general point regarding construction of the index is that erring on the kernel size (relative to the "true" bandwidth at which the spatial relationship between churches and congregations is determined) would only weaken the instrument in the first stage. Our main results present multiple bandwidths to confirm that our findings are not sensitive to the bandwidth size.

Table 3.O.1: Robustness of IV Tobit Estimates to Mixing Kernel Density

	First Stage *Dependent variable: Diversity*
	(i)
Kernel Bandwidth:	Mixed
Church Diversity	0.28***
	(0.01)
ln(Population)	-0.08***
	(0.01)
ln(Population)^2	0.01***
	(0.00)
N	6,838
R^2	0.07
F-Statistic on Instrument	454.83

	Second Stage *Dependent variable: Number of Victims*
	(i)
Kernel Bandwidth:	Mixed
Ethnic Diversity	10.26***
	(1.28)
ln(Population)	-2.38***
	(0.33)
ln(Population)^2	0.48***
	(0.05)
N	6,838
Wald χ^2	309.42

**p<0.05; ***p<0.01

Regressions include a constant, not shown.

Church diversity constructed using 500m kernels for Protestant churches and 1 km kernels for Catholic churches.

3.P Changing Unit Size

We test the sensitivity of our results to the spatial resolution of the data by aggregating the original 100 meter cells to 500 meter cells. We estimate the Tobit specification using the 1 kilometer kernel density for churches and present the results in Table 3.P.1. The unadjusted results, shown in the first column, show that diversity is still strongly associated with the number of deaths. The instrument remains strong in the first stage, with an F-statistic of 105.9. The coefficient of 0.29 is qualitatively unchanged from the coefficient of 0.33 using the original 100 meter grid cells (see Table 3.2 in the paper). The second stage results corroborate our original finding, that diversity is strongly associated with more violence in a cell. The second stage results corroborate our original finding, that diversity is strongly associated with more violence in a cell (the coefficient is predictably larger than when using 100m cells, given that the cell sizes are larger and our dependent variable is not defined per unit area). As in our original model, the results using 500 meter cells show a significant attenuation bias when using unadjusted data to estimate diversity's effect. The coefficient in the IV Tobit specification (26.39) is almost double the magnitude estimated by the Tobit model using unadjusted census data (14.96).

Table 3.P.1: Robustness of IV Tobit Estimates to Changing Unit Size

	Dependent variable: Diversity	
	(i)	(ii)
Estimation Strategy:		First Stage
Church Diversity		0.29^{***}
		(0.03)
ln(Population)		-0.04
		(0.02)
ln(Population)^2		0.00
		(0.00)
N		577
R^2		0.16
F-Statistic on Instrument		105.90
	Dependent variable: Number of Victims	
	(i)	(ii)
Estimation Strategy:	Tobit	IV Tobit – 2SLS
Ethnic Diversity	14.96^{***}	26.39^{***}
	(2.15)	(5.70)
ln(Population)	-7.62^{***}	-7.23^{***}
	(1.25)	(1.28)
ln(Population)^2	1.07^{***}	1.03^{***}
	(0.12)	(0.12)
N	577	577
Wald χ^2		349.31

$^{**}p<0.05$; $^{***}p<0.01$

Regressions include a constant, not shown.

3.Q Non-Linear Effects

To check for nonlinearities in the relationship between diversity and violence, we separate both the diversity variable and the instrumented diversity variable (the fitted values from the first-stage regression) into quintiles. We then regress the number of deaths on dummies for the quintiles. The fitted values of deaths by diversity quintile (at the mean level of population) using the raw diversity variable from the census are as follows:

Table 3.Q.1: Fitted Values using Quintiles of Diversity

	Dependent variable: Predicted Number of Victims			
Quintile	(i)	(ii)	(iii)	(iv)
1	0.12	0.15	0.09	0.06
	(0.09, 0.14)	(0.10, 0.19)	(0.06, 0.12)	(0.03, 0.08)
2	0.18	0.09	0.06	0.03
	(0.14, 0.22)	(0.04, 0.13)	(0.02, 0.09)	(0.01, 0.06)
3	0.31	0.18	0.12	0.06
	(0.26, 0.36)	(0.13, 0.22)	(0.09, 0.15)	(0.04, 0.08)
4	0.32	0.28	0.19	0.09
	(0.26, 0.39)	(0.23, 0.32)	(0.16,0.22)	(0.07, 0.11)
5	0.49	0.32	0.19	0.14
	(0.40, 0.57)	(0.29, 0.37)	(0.16, 0.22)	(0.11, 0.16)

95% confidence intervals in parentheses.
Column (i) uses the unadjusted diversity variable.
Column (ii) uses the instrumented diversity variable.
Column (iii) uses the instrumented diversity variable subset on civilian victims.
Column (iv) uses the instrumented diversity variable subset on non-civilian victims.

We note from these predicted values that the amount of violence is monotonically increasing in diversity (by unadjusted estimates) and that the highest amount of violence occurs in the highest diversity quintile. We observe a similar trend in the predicted values of deaths by quintile of the instrumented diversity variable (Column ii). With the exception of the least diverse areas (quintiles 1 and 2) violence is monotonically increasing in diversity. The highest levels of violence occur in the uppermost quintile of the instrumented diversity variable. Similar trends are observed in Column (iii) using only civilian deaths and Column (iv) using only non-civilian deaths.

These results suggest that neither our unadjusted nor our 2SLS results are driven by any particular part of the distribution of the diversity variable.

4 Social Activity Hubs: Estimating User Specific Contextual Factors from Social Media Data

Abstract

Context influences sociopolitical attitudes and behaviors, making the estimation of individuals' contexts an important methodological problem for the social sciences. We add to this body of work by presenting a method to estimate an individual's spatial contexts, specifically the set of geospatial areas an individual is most active in. Our approach, which utilizes the Dirichlet process mixture model, departs most significantly from more traditional approaches to estimating relevant spatial locations in that it does not arbitrarily constrain the number of spatial contexts an individual can have. This modeling approach reflects our recognition that an individual's lived experiences is a combination of different contexts that overlap to varying degrees. This flexibility therefore yields a more valid measure of spatial contexts. To illustrate our method, including its performance relative to other measures, we apply our method to Twitter data generated by protesters who participated in the 2015 Freddie Gray protests in Baltimore, MD.

4.1 Introduction

Across a wide variety of social sciences, context has been repeatedly shown to be one of the most important determinants of human social behavior. Context matters. Individuals modulate their responses to stimuli according to their social network, the formality of the situation, and power and status differentials. While qualitative accounts of behavior have long emphasized the importance of context, recent empirical research provides causal evidence for the effect of context on behavior. Using a variety of sophisticated research designs, context has been shown to affect altruism (Sands 2017); anti-immigrant sentiment and discriminatory behavior (Enos and Gidron 2016; Enos 2016, 2014); and support for extremist politicians (Getmansky and Zeitzoff 2014).

The relationship between context and behavior in the social sciences is not always easy to discern. Each individual's behavioral responses are conditioned by the social, economic, and political realities of their specific setting. Unfortunately, identification of individual-level contextual factors that are hypothesized to influence behavior are difficult to gather. Furthermore, it is often the case that the act of measurement itself changes the behavior of the subject under study. Measurement is typically an inherently social act. Survey participation requires the active and informed consent of the individual participant. In this context, social desirability bias is quite difficult to avoid. Thus, contextual factors at the individual level are difficult and costly to measure and often too complex to manipulate experimentally.

To overcome these challenges, many of the aforementioned studies rely upon some form of randomized intervention as a design feature, be it through the lab, in the field, or naturally occurring. In these studies, each subject's context is manipulated through various stimuli. For example, in one study participants encountered a homeless person, staged by the experimenters, as they completed a survey. This allowed the experimenters to measure the way in which visible poverty affects altruistic sentiment (Sands 2017). Another study engaged participants who self-identified as white/Caucasian in a game in which they were asked to play the role of a country's dictator. When these participants had been previously informed about the growth of the Hispanic population inside the United States, they were more likely to favor other white participants (Abascal 2015). Research has also found that transphobia — prejudice against transgender people — is reduced when canvassers randomly encourage active-perspective-taking among survey respondents (Broockman and Kalla 2016). These studies collectively illustrate the powerful way in which situational factors can influence social behavior.

Randomized experiments convincingly establish causality and rule out alternate explanations. However, the external validity of such findings are not always easily established (Fariss and Jones 2017; Shadish 2010; Shadish, Cook, and Campbell 2001). This problem is particularly acute in experiments that attempt to measure political and social context, which, by necessity, tend to adopt extremely powerful treatments in order to generate statistically detectable effects within limited sample sizes. In contrast, though observational studies do not typically have as strong internal validity, they effectively address the question of external validity. Observational studies measure subject specific-behaviors in real life situations, often without direct intervention on behalf of the experimenters, which may unintentionally alter the behavior under study (Fariss and Jones 2017). For example, observational studies of political context have considered the way in which demographics affect "white flight" (Crowder 2000; Kruse 2005), diversity and inter-ethnic violence (Brown, McCord, and Zachary 2017), and voter turnout for minority candidates (Barreto, Segura, and Woods 2004).

Most theories about behavioral variations across contexts operate at the individual level, but inference at this level of analysis is often constrained by two primary difficulties associated with data availability. First, sparsity in behavioral responses measured at the individual level often leads to the use of group-level outcomes as aggregated behavioral measures. The resulting inference may be subject to a serious ecological inference problem (King 2013). Second, subject to data constraints, researchers are often forced to rely on coarse measures of context. Depending on the particular mechanisms specified in the study, the use of coarse, non-localized contexts may be inappropriate as there might be high levels of within-context variation in the stimuli of interest. Imprecise measures of context therefore introduce high levels of measurement error, rendering the inferred contextual factors not meaningfully relevant to individuals' experiences. Scholars have adopted a variety of strategies to address this issue of ecological validity, such as using small units of analysis when available (Brown, McCord, and Zachary 2017), or adopting different statistical methods. The increasingly widespread penetration of social media, and the resulting flood of individual-level data, presents a third option: measuring context through social media behavior (Lazer et al. 2009; Lazer and Radford 2017).

An increasingly rich and important literature uses social media to draw inferences about contextual effects across a variety of domains and behaviors, including job-seeking (Gee et al. 2017), voting behavior (Bond et al. 2012), collective action in autocracies (Steinert-Threlkeld 2017; Steinert-Threlkeld et al. 2015), and even politicians' communication with one another (Barberá 2014). Commonly-available forms of social media data still pose restrictions on the types of contextual effects that researchers can study. Although the application program interface (API) for some companies and commercial vendors provides information about users' time zone or country, these are measures of context only in the broadest and most aggregate sense. Using these locations to infer context potentially creates the same type of ecological inference issue as described above. Again, it is crucial to address this issue, as we know that an individual's context matters for understanding how she conditions her behaviors.

To address this inferential problem, there is a burgeoning field of research that develops new methods for analyzing what is referred to as "volunteered geographic information" (Jiang and Thill 2015). Grace et al. (2017), for example, uses users' network ties to local organizations to infer residential location. Others, such as Hasan, Zhan, and Ukkusuri (2013), employ semantic analysis to extract geospatial information embedded in social media text. This field of research highlights the importance of continually developing means to infer context from social media data.

In this paper, we contribute to this developing line of research and present a method to estimate Twitter users' "social activity hubs" (SAHs), or the geospatial areas where users spend their time. Our approach, which builds on Rossmo (1999) and Verity et al. (2014), departs most significantly from earlier approaches in that it does not arbitrarily constrain the number of clusters an individual's overall movement profile can contain (Verity et al. 2014). This modeling approach reflects our recognition that an individual's lived experiences is a combination of different contexts that overlap to varying degrees, and that data-driven methods of inferring how many contexts are relevant to each individual is superior to relying on assumptions when we lack strong *a priori* beliefs. We estimate these SAHs using an algorithm that selects between geoprofiling models of varying sophistication that are conditional on information availability. For users with enough information, we utilize posterior quantities from a Dirichlet process mixture model to compute SAHs.

As initial evidence the utility of SAHs as a measure of political context, we run our algorithm on a sample of geotagged tweets made by Twitter users who participated in the Freddie Gray protests in Baltimore, MD in April 2015. Estimated SAHs are plotted on a Baltimore City map in Figure 4.1. As is evident, our model yield SAHs that cluster in areas with high daytime populations, such downtown Baltimore, and the high population area surrounding Johns Hopkins University. This demonstrates that our SAHs are a useful tool to generate disaggregated, individual-level estimates of social media users' economic, social, or political contexts. In the remainder of this paper, we describe the SAH model, and further illustrate its applications using Twitter data from Baltimore during the Freddie Gray protests.

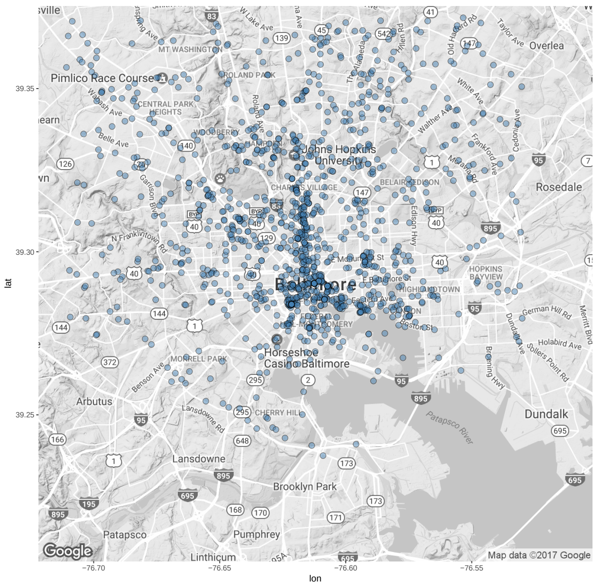

Figure 4.1: Geographical distribution of social activity hubs in Baltimore, MD estimated using the DPM-based local minima submodel. SAHs cluster in areas with high daytime populations like downtown Baltimore. The estimation process is described in section 4.3.

4.2 Research Design

4.2.1 Background on the Social and Political Context during the 2015 Baltimore Protests

Freddie Gray was arrested on April 12, 2105 by Baltimore Police Department (BPD) officers for possession of what officers believed at the time was an illegal switchblade.[1] For reasons that are under dispute, Gray fell into a coma while being transported in a police van subsequent to his arrest. He never recovered from his injuries and died in a trauma center on April 19, 2015. Starting on April 18, 2015, protesters began gathering in front of Baltimore's Western district police department to denounce Gray's alleged mistreatment and BPD brutality. These protests grew steadily in size as media attention to Gray's case increased throughout the

[1] While officers testified that they believed Gray's knife was illegal, the Maryland state attorney for Baltimore later clarified that Gray in fact was in possession of a spring-assisted knife that was legal under Maryland law (Blinder and na 2015).

week. The protests continued to gain momentum and eventually reached several thousand people. The Maryland National Guard, responding to a declared state of emergency, was brought in to restore order to the city. A mandatory curfew was declared within Baltimore city limits from April 28 to May 3. As Chen, Zachary, and Fariss (2017) argue that contact with police affects behavior, our empirical application is focused on these protests.

4.2.2 Sampling and Data Collection

Prior to beginning our analysis, we first purchased all of the geotagged tweets posted within the geospatial boundaries of Baltimore City, MD from April 16, 2015 to May 4, 2015. These dates were selected because they were when the protests related to the death of Freddie Gray occurred in the city. Our sample included a total of 111,440 tweets made by 7,884 unique users. In order to restrict our analysis to people with a positive probability of protesting, we limited our sample to only include geotagged tweets. Present illustrations are based on smaller subsets of these users.

In order to estimate these users' SAHs, we collected up to 3,200 of each user's most recent tweets using Twitter's API, which we called through the TwitteR package in R (Gentry 2015).[2] Tweets were collected between July 19, 2016 and August 27, 2016. Each user's tweets are then narrowed to those that contain geotags. Each geotagged tweet in this final sample is treated as an observed incident of the user's movement patterns, and from the collection of all observed incidents, we estimate the user's SAHs.[3]

By default, a twitter user's location is not displayed when posting a 140 character message

[2] API Documentation available here: `https://dev.twitter.com/rest/reference/get/statuses/user_timeline`.

[3] As our research strategy enables us to estimate Twitter users' Social Activity Hub location, it is pertinent to address ethical concerns regarding the steps we take to maintain anonymity and protect users from potential harm. First, the sample is only taken from users who had opted into sharing their location with Twitter. By default, Twitter does not record the location where a tweet was posted. Instead, users must change their phone's settings to give Twitter permission to record their location via GPS. Second, we anonymize Twitter account names by applying a cryptographical hash. Third, the estimate standard deviation of the location means that we are only able to know the location of the Social Activity Hub within three miles.

to twitter. However, users can identify their location when tweeting by enabling the location services that twitter provides. A user is able to selectively add location information, such as a geographic area (city or neighborhood), or a precise location in terms of latitude and longitude coordinates from the global positioning system that is available in most smart phones. Importantly, opting to share the location of a tweet is a social act. Because this is central to our research question, we do not include data from those individuals who chose not to disclose their geographic location.

4.3 Estimating Social Activity Hubs

In this section, we present in detail the method we used to estimate Twitter users' SAHs. As the availability of information associated with each Twitter account differs, our SAH model, summarized in algorithm 3, is conditional on what this information affords, defaulting to more basic models where data availability is low. More specifically, we intend to define SAHs in two ways described in more detail below, based on posterior quantities of a Dirichlet process mixture (DPM) model for spatial data (Verity et al. 2014). Estimation relies on an MCMC algorithm,[4] which is subject to convergence difficulties. In these rare cases, we document the specific user and return to diagnose potential issues. We discuss this phenomenon and our solutions more explicitly in section 4.3.1.

4.3.1 Dirichlet Process Mixture Model for Spatial Data

For users whose tweets contain sufficient information regarding their movement patterns, we use the Dirichlet process mixture (DPM) model of geographic profiling as the basis of our SAH model. DPM models for spatial data, based on prior geographic profiling models in criminology (O'Leary 2010; Rossmo 1999), was first described in (Verity et al. 2014) where it was applied to

[4] The algorithm is implemented in the Rgeoprofile package for R (Stevenson et al. 2014).

Algorithm 3: Social Activity Hub Estimation for Each User

```
Data: The set of n observed incidents x_i = (x_i^(1), x_i^(2)), i = 1,...,n
if n = 1 then
    Assign x_1 as SAH^(sole observed);
else
    MCMC algorithm implemented as the RunMCMC() function in Rgeoprofile 1.2
      (Stevenson et al. 2014: summarized below in algorithm 4), based on the
      discussion in Verity et al. (2014));
    if convergence fails;
     then
        Document failure;
    else
        Take 3000 posterior draws; thin by keeping the first of every 30;
        begin local minima model:
            combine all 100 posterior draws;
            a) calculate hitscore surface;
            b) find local minima j on surface within σ degree decimal radius,
              j = 1,...,∞;
            c) foreach local minimum j do
                Assign x_i closest in Euclidean distance to local minimum as
                  SAH_j^(local minima)
            end
            foreach posterior draw do
                a to c;
            end
        end
        begin cluster mean model:
            combine all 100 posterior draws;
            d) foreach cluster j of x do
                Assign x_i closest in Euclidean distance to estimated source of cluster
                  as SAH_j^(cluster mean)
            end
            foreach posterior draw do
                d;
            end
        end
    end
end
Result: SAH=(SAH^(sole observed), SAH^(local minima), SAH^(cluster mean))
```

spatial epidemiology. More recently, the model was used in an attempt to determine the identity of graffiti artist Banksy (Hauge et al. 2016).

The intuition of the DPM model for spatial data is to sort a set of observed incidents in physical space into clusters originating from different source locations, without prior assumptions about the number of clusters that exist. For our present purposes, the DPM model is preferred over alternatives that require a fixed number of clusters (including those with a single cluster), because individuals are likely to vary in terms of their movement patterns (of which we have no prior data). Where there are multiple clusters, especially when they are highly dispersed, a misspecified number will result in inaccurate source location estimates that are skewed by "outliers," which are actually observations that originate from a different source.

The DPM model rectifies this by estimating the number of sources based on the observed data. The flexibility afforded by this feature is especially desirable, given the large number of Twitter accounts we are working with, as it is not feasible to adjust the SAH model for each Twitter account individually. A DPM model is not without assumptions, which are provided in the description below. In short, by employing the DPM model, we assume that individuals can have multiple SAHs from where their movement outward follows identical distributions, which in this implementation we specify as a bivariate normal distribution, with standard deviation varying by specific application.

More specifically, the DPM model we use, adapted for spatial data by Verity et al. (2014), is as follows. For each Twitter user, define a two-dimensional sample space with a finite grid of cells as Ω, in which each cell $\boldsymbol{\omega} = (\omega^{(1)}, \omega^{(2)})$ is a vector containing the latitude and longitude in decimal degrees of a geocoordinate. The set of n geocoordinates obtained from geotagged tweets $\mathbf{x} = \mathbf{x}_{1,\ldots,n}$ is assumed to be the result of independent draws from a mixture of a countably infinite set of bivariate normal distributions centered on $\mathbf{z} = \mathbf{z}_{1,\ldots,\infty}$, each with a variance of σ^2; (σ contains expectations about the movement patterns of individuals and must be specified by the user). Both $\mathbf{x}$ and $\mathbf{z}$ are defined on Ω. The prior distribution of the set of $\mathbf{z}$ is assumed to be a bivariate normal centered on the mean of $\mathbf{x}$, with a variance of τ^2 (τ is set to the largest distance in either longitude or latitude). c_i is a categorical variable that assigns $\mathbf{x}_i$ to source $\mathbf{z}_{c_i}$, and is drawn from a Dirichlet process, specifically the Chinese Restaurant Process which has a concentration parameter α drawn from a diffuse hyper-prior (specifically $h(\alpha) = ((1+\alpha)^2)^{-1}$) and a base distribution that is the bivariate normal (with mean $\mathbf{x}/n$) discussed above. This is formally represented as,

$$\begin{aligned} \mathbf{x}_i|\mathbf{z}_{c_i} &\sim \mathcal{N}(\mathbf{z}_{c_i}, \Sigma = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}) \\ \mathbf{z}_{1,\ldots,\infty} &\sim \mathcal{N}(\mathbf{x}/n, \mathrm{T} = \begin{bmatrix} \tau^2 & 0 \\ 0 & \tau^2 \end{bmatrix}) \\ c_i &\sim CRP(\alpha) \\ \alpha &\sim \mathcal{H} \end{aligned} \tag{4.1}$$

Exact computation of posterior quantities are intractable when the number of observations is high ($n > 10$ being a useful rule of thumb; see (Verity et al. 2014) for analytical solutions to relevant posterior quantities), but can be estimated using MCMC methods (Neal 2000; Verity et al. 2014), which is implemented in the R package Rgeoprofile 1.2 (Stevenson et al. 2014). The MCMC algorithm (`RunMCMC()` presented in algorithm 4) is based on a Gibbs sampler that alternates between draws of source locations $\mathbf{z}_{c_i}$ and cluster assignment c_i for all $i = 1, \ldots, n$ observations. The algorithm returns, for each $\mathbf{x}_i$, its cluster c_i; and for each unique cluster c_j, its spatial mean $\mathbf{z}_j$.

Diagnosing Convergence Difficulties

MCMC convergence is assessed by the potential scale reduction factor (psrf) evaluated on the log-likelihood of the model (Gelman and Rubin 1992). This assessment is implemented as `gelman.diag()` the in the coda package. MCMC chains are taken to have converged when the

Algorithm 4: `RunMCMC` from Rgeoprofile 1.2 (Stevenson et al. 2014)

Data: The set of n observed incidents $\mathbf{x}_i = (x_i^{(1)}, x_i^{(2)})$, $i = 1, \ldots, n$
Initialize by setting initial values and computing relevant priors;
Define sampling steps:
  a) draw and update $\mathbf{z}_{c_i}$ based on most updated c_i;
  b) draw and update c_i based on most updated $\mathbf{z}$;
begin Burn-in
  repeat
    for *i in* 1 *to n* **do** a-b;
  until *convergence*;
end
begin Posterior draws
  foreach *posterior draw* **do**
    for *i in* 1 *to n* **do** a-b;
  end
end
Result:
  1. For each $\mathbf{x}_{1,\ldots,n}$, its corresponding cluster c_i
  2. For each unique cluster c_j, its source location $\mathbf{z}_j$

upper bound of the psrf falls below 1.1 following a burn-in period, which we specified as 300 draws. Generally, models successfully converge within the burn-in period or shortly after, but one of two issues may arise.

First, a small number of models take a disproportionally long time to reach convergence. This usually occurs for users with a large set of observed incidents. Because we have a need to estimate a large number of SAHs, we specify a maximum burn-in of 3,000 iterations, at which point if the psrf is not below 1.1, the specific user is documented for manual diagnosis. For reference, consider that in the two illustrations presented below, failure to converge after 3,000 burn-in draws occurred 14 times out of 200 users estimated, and one time out of 126 users estimated. In fact, for most data sets, convergence was achieved within the minimum burn-in period of 300 or shortly after. For the present illustrations, we drop users who do not converge after 3,000 burn-in draws from our examination as representative sampling is not a requirement.

Second, extreme sparsity in data in the form of singular observations or spatially dispersed observations without overlap (which are singular observations within certain clusters) can result in an error when computing the psrf. To see why this occurs, first note that the log-likelihood of the model is calculated based on the fit of the distribution of observed incidents into clusters, including the number of clusters present and how the observed incidents are sorted among them. This occurs after sampling step b in algorithm 4. In instances of high data sparsity and dispersion, cluster assignments c_i are never updated through sampling step b because the probability of assigning a different cluster, which is updated in step a, while always nonzero, is extremely low. The result is that for all MCMC iterations across all chains, the same log-likelihood is computed based on the unchanging distribution of observed incidents into clusters. In short, the Gibbs sampler immediately moves to a very small area and any movement within this area does not yield probabilities for different clustering combinations that is meaningfully above zero. The log-likelihood which is computed based on this clustering therefore remains constant, leading to errors when attempting to compute the psrf as it is based on variation within and across MCMC chains. We document these errors, but take the modeled results as the best estimate of SAHs given the available information, and as such, take these models as having converged. The most extreme case is where there is only a singular observation, which we immediately take to be the SAH as outlined in algorithm 3.

4.3.2 Local Minima and Cluster Mean Submodels

As introduced earlier, we use the posterior quantities obtained from the DPM model in our SAH model in two ways. For the *local minima* submodel, begin by defining $\mathbf{S} \subseteq \Omega$ as the grid bound by the minimum and maximum values of the set of observed $\mathbf{x}$. Next, for every cell $\mathbf{s} \in \mathbf{S}$, rank $\mathbf{s}$ according to the sum of its distances to each source location $\mathbf{z}_j$ over all posterior draws, where distance is not linear but weighed by the inverse of the bivariate normal density around $\mathbf{z}_j$. Consistent with existing geoprofiling approaches (e.g. Rossmo 1999), ranks are transformed to hitscores on $[0,1)$, but remain functionally equivalent in that lower is better and all values are distinct.[5] This type of hitscore surface is traditionally used as a surface for search priority of source locations (Rossmo 1999; Verity et al. 2014). On this surface, we find all m local minima (i.e. locations with higher priority) within an approximately two mile radius (0.05 decimal degrees) and define a user's SAHs as the set of m observed $\mathbf{x}_i$ closest to these local minima. For the *cluster mean* submodel, we define a user's SAHs as the set of observed $\mathbf{x}_i$ closest to the set of estimated source locations $\mathbf{z}$ averaged across posterior draws.

Figure 4.1 illustrates the SAHs estimated under both submodels in relation to the hitscore surface produced by the DPM model. For this Twitter account, the DPM model aggregated over all posterior draws estimated the set of observed incidents to have originated from two sources (i.e. $\mathbf{z}_j, j = 1, 2$). As evident from Figure 4.1, the two submodels agreed on a potential source $\mathbf{z}_1$ in the upper left of the physical space (directly north of Baltimore City) as an SAH. In the bottom left (directly west of Baltimore City), observations are not dispersed enough to consistently yield a third cluster, but are relatively sparse, such that the estimates for the second source varied greatly. In fact, between different posterior draws, the DPM model assigned the set of $\mathbf{x}$ not associated with $\mathbf{z}_1$ to either $\mathbf{z}_2$ or $\mathbf{z}_3$. Because of this, the difference between the local minima and cluster mean submodels (based on how posterior draws are aggregated, i.e., sum of computed probabilities versus means), leads to disagreement between the submodels on the second SAH. This example illustrates the importance of understanding uncertainty in the DPM model, which we discuss next.

4.3.3 Uncertainty in the DPM Model

In earlier applications of the DPM model to spatial data, there is justifiably less of a concern over the uncertainty of estimates. However uncertain, the expected values of $\mathbf{z}$ are what informs a search that must take place. Existing implementations of the model (Stevenson et al.

[5] The two computational steps above are implemented in the `ThinandAnalyse()` function in Rgeoprofile 1.2 (Stevenson et al. 2014).

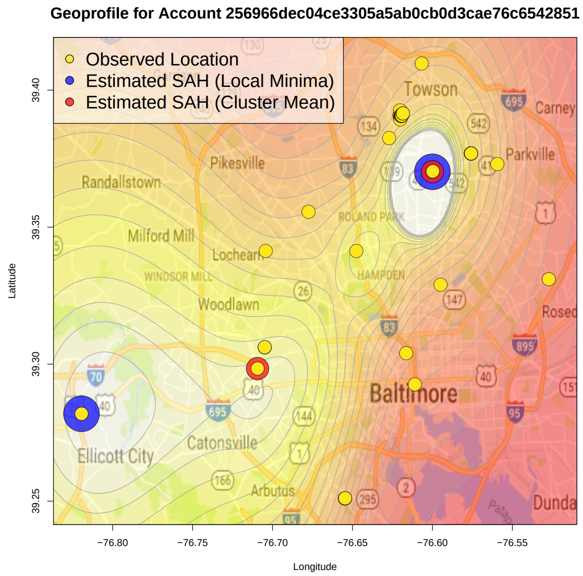

Figure 4.1: Example of a single user's estimated social activity hubs in relation to the hitscore surface produced by the DPM model. Yellow points are observed incidents. Points in blue are determined by the local minima submodel. Points in red are determined by the cluster means submodel.

2014: e.g.,) therefore do not readily yield uncertainty measures. However for inferential modeling, measures of uncertainty feature much more prominently. In order to account for uncertainty in our SAH model, we take 3,000 draws from the posterior distribution of the DPM model, thinned to 100 samples, and use this information to determine a set of corresponding SAHs following both the local minima and cluster mean submodels. Specifically, for the local minima submodel, instead of computing a hitscore surface based on all posterior draws, we do so for each draw independently; and for the cluster mean submodel, $\mathbf{z}$ is not averaged across posterior draws. SAH estimates are stored for each posterior draw, forming a posterior distribution of SAHs. This distribution can be used in subsequent statistical modeling to account for uncertainty associated with the SAH model. In the remainder of this section, we use the same Twitter account as above to illustrate uncertainty within the two DPM-based submodels. This particular account was chosen because it is illustrative both in terms of its estimates and the uncertainty associated with them. The level of uncertainty associated with this account, based on visual inspection, is neither particularly high nor low.

In order to visualize the level of uncertainty about our SAH estimates, we plot the variation in hitscore associated with each potential source across all posterior draws. Specifically, in Figure 4.2 (which corresponds to the hitscore surface in Figure 4.1), each horizontal line documents the hitscore of a potential source location as it varies across posterior draws. The highlighted lines are the sources chosen as the SAHs from the combined posterior draws, which may differ depending on the submodel used. The highlighting color serves to tie corresponding SAHs across the two subfigures. Variation in the hitscore indicates changes in the topography of the hitscore map across posterior draws. This does not, however, necessarily mean that there is uncertainty about the SAH estimates, which arise when changes in the topography are large enough to induce changes in the hitscore rank of potential sources relative to one other. This uncertainty is indicated by lines that cross. Future work may benefit from a formal quantification of this type of uncertainty.

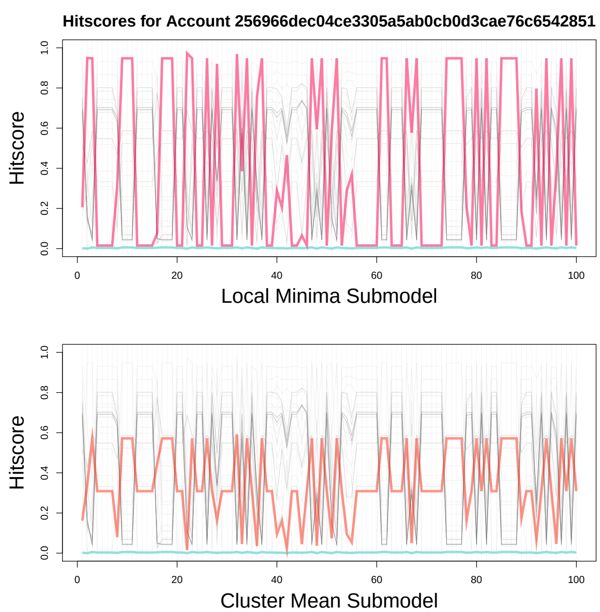

Figure 4.2: Visualization of uncertainty about SAH estimates. Each line indicates a source location and its variation in hitscore across different draws. Highlighted lines indicate source locations selected by the algorithm as an SAH.

As discussed earlier, the DPM model estimated the set of observed incidents associated with this account to have originated from two sources. As is shown in Figure 4.2, corroborating Figure 4.1, both DPM-based submodels yielded the potential source location with the lowest hitscore as one of the two SAHs. This SAH_1, highlighted in turquoise, is associated with very little uncertainty. The variation in its hitscore across all posterior draws is minor, and it maintains a stable hitscore rank in all but one draw (42). On the other hand, the second SAH_2, highlighted in two shades of coral, differs between the two submodels and exhibits high levels of uncertainty within each submodel. As is evident from the figure, the local minima $\mathrm{SAH}^{(localminima)}$ varies widely between posterior draws, but for the majority of these draws, it is the best performing SAH (in terms of hitscore), just after the SAH_1. The cluster mean $\mathrm{SAH}^{(clustermean)}$ is not selected based on the hitscore of potential source locations, but still yields a relatively desirable result according to this criterion; while its hitscore based on the aggregated posterior draws is not as low as that of the $\mathrm{SAH}_2^{(localminima)}$, it is subject to less of the extreme fluctuation across draws exhibited by the local minima submodel.

At the same time, the present discussion serves to illustrate the potential for high levels of uncertainty to be associated with DPM-based SAHs. Researchers intending to utilize these estimated quantities should be mindful of this when making statistical inferences. As we proposed, the entire posterior draw can be used in statistical modeling, which allows for the construction of confidence bounds. Our novel method paves the way future work in formal quantification of the uncertainty discussed here and in better understanding the inferential shortcomings associated with not properly accounting for this source of uncertainty.

4.4 Illustrations

In this section, we illustrate the validity of our measurement algorithm in two ways, as comparison to traditional spatial estimation models, and based on its predictive validity.

4.4.1 Comparison to Spatial Means

First, we compare for each user the performance of their SAHs obtained from the local minima and cluster mean submodels relative to the source closest to the spatial mean of their observed incidents. To do so, we randomly selected 200 users from the sample described in section 4.2.2. Users whose DPM model did not converge were dropped as discussed in section 4.3.1, yielding a final sample of 186. For each user, we estimated their $\text{SAH}^{(localminima)}$ and $\text{SAH}^{(clustermean)}$.[6] We also calculated the spatial mean of their observed incidents, and similar to how we estimate SAHs, specify the potential source closest to the spatial mean as the estimated source. Then, we calculate the Euclidean distance between each of the user's observed incidents to its closest $\text{SAH}^{(localminima)}$, $\text{SAH}^{(clustermean)}$, and the spatial mean-estimated source. Finally, we record the median distance for each estimation method. We repeated these steps for each of the 186 users. The results are summarized in Table 4.1 as percentiles of the median distances for each estimation method.

Table 4.1: Comparison of different estimation methods. Cells are the percentile values of the median distance in decimal degrees from each user's observed incidents to their closest SAH/spatial mean.

Distance between Estimated Sources and Incidents

	Percentiles				
Method	5	25	50	75	95
Local Minima	0.00	0.0005	0.02	0.07	0.40
Cluster Mean	0.00	0.0004	0.01	0.02	0.05
Spatial Mean	0.00	0.0651	0.33	1.67	20.22

$n = 186$

[6] Model parameters are specified to be the ones presented: $\sigma = 0.05$ decimal degree; minimum burn-in of 300; maximum burn-in of 3,000; 3,000 posterior draws thinned to 100. Convergence results are: 14 did not converge after 3,000 burn-in draws. Data for 42 users were at least conditionally sparse in observations, with 9 of these being single observations.

It is evident from the results that the source estimated based on the spatial mean tends to be considerably more distant to the user's observed incidents. Based on this, we believe that the face validity of the spatial mean is generally low. These results highlight the importance of a data-driven approach to estimating the optimal number of clusters in social media data. Forcing data into a user-determined number of clusters (one in this example) biases estimated movement patterns, which may critically influence subsequent inferential steps.

4.4.2 Predictive Validity

Next, we assess the predictive validity of our SAH measures. To do so, we compare whether the SAHs from known protesters in our data are more likely to fall within known protest locations compared to a sample of users whose protest behavior is unknown.

After drawing a random subset of 1,000 tweets from the sample described in section 4.2.2, we hand-code whether each tweet indicated attendance in the Freddie Gray protests. We identify 64 unique protesters within this subset. From the same subset, we randomly select 62 users who we did not identify as protesters. Then, we estimate our two SAH measures for all 126 users.[7]

Using the resulting SAHs, we next consider whether there are different geospatial patterns in the SAHs from protesters and non-protesters. Specifically, we assess the proportion of users from both samples with at least one SAH within a known Freddie Gray protest location. We identify these locations based on data from Baltimore news sources published in April and May, 2015. These locations are summarized in Table 4.2.

Results from our examination are presented in Table 4.3. As these results indicate, there is a stark contrast between the SAH of users from whom we have observed protest behavior and those for whom we have no information. These results demonstrate the ability of our measure

[7] Because the aim is to identify localized movement patterns as opposed to more general hubs, model parameters are specified slightly differently: $\sigma = 0.01$ decimal degree; minimum burn-in of 300; maximum burn-in of 3,000; 3,000 posterior draws thinned to 100. Convergence results are: 1 did not converge after 3,000 burn-in draws. Data for 24 users were at least conditionally sparse in observations, with 10 of these being single observations.

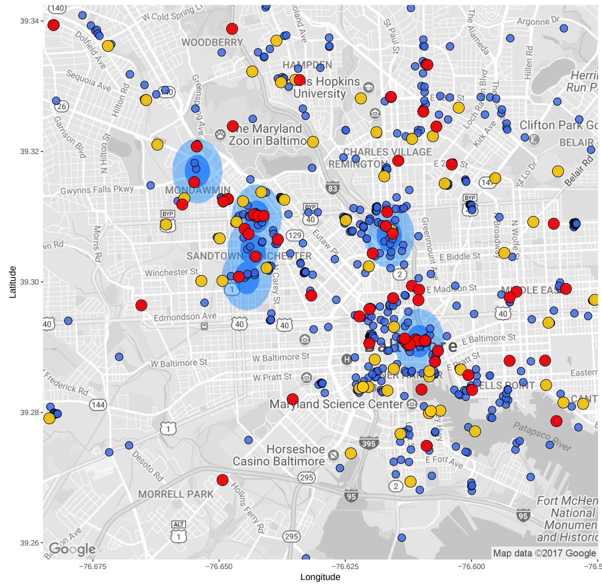

Figure 4.1: The six light blue circles indicate known protest locations. Dark blue dots are all observed tweets. Red dots are SAHs that belong to protesters, while yellow ones belong to non-protesters. These results are presented in Table 4.3.

Table 4.2: Summary of locations in or near Baltimore City bounds specified as a protest location during the Freddie Gray protests, April-May, 2015.

Locations of Freddie Gray Protests

Location	Longitude	Latitude
North and Penn	-76.6425	39.3100
Baltimore Police Dept. West	-76.6445	39.3006
Baltimore City Hall	-76.6104	39.2909
Gilmor House	-76.6433	39.3049
Mondawmin Mall	-76.6543	39.3170
Penn Station	-76.6163	39.3071

to capture meaningful behavior patterns based on observed movements, further supporting the validity of our method.

Table 4.3: Summary of SAH predictive validity. Cells are the proportion of users from either sample with at least one estimated SAH falling within a known protest location. Protest location is defined by a 0.0025 decimal degree radius around each of the six coordinates specified in Table 4.2.

Proportion of Users with SAHs in Protest Locations

	Local Minima	Cluster Mean
Protesters	0.37	0.27
Random Sample	0.02	0.02
Difference	0.35	0.25

$n_{protest} = 63$, $n_{random} = 62$

4.5 Discussion

In this paper, we contribute to the developing field of research on inferring geospatial context from volunteered geographic information, and presented a method to estimate Twitter users' "social activity hubs" (SAHs), or the geospatial areas where users spend time throughout the day. As a validation exercise, we linked these locations to incidences of political participation, in particular the protests that transpired over the death of Freddie Gray, in Baltimore during April and May of 2015. The patterns we discovered suggest the methods proposed here for estimating SAHs are able capture meaningful measures of sociopolitical context.

Chapter 4, in full, is a reprint of the material as it appears in Chen, Ted Hsuan Yun, Paul Zachary, and Christopher J. Fariss. Computational Social Science 2017: Proc. 2017 Annual Conf, David Krakauer and Scott Page, Eds.

Bibliography

Abascal, Maria. 2015. "Us and them: Black-White relations in the wake of Hispanic population growth." *American Sociological Review* 80 (4): 789–813.

Barberá, Pablo. 2014. "Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data." *Political Analysis* 23 (1): 76–91.

Barreto, Matt A, Gary M Segura, and Nathan D Woods. 2004. "The mobilizing effect of majority–minority districts on Latino turnout." *American Political Science Review* 98 (1): 65–75.

Blinder, Alan, and Richard Pérez-Pe na. 2015. "6 Baltimore Police Officers Charged in Freddie Gray Death." *New York Times* .

Bond, Robert M., Christopher J. Fariss, Jason J. Jones, Adam D.I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. 2012. "A 61-Million-Person Experiment in Social Influence and Political Mobilization." *Nature* 489 (7415): 295–298.

Broockman, David, and Joshua Kalla. 2016. "Durably reducing transphobia: A field experiment on door-to-door canvassing." *Science* 352 (6282): 220–224.

Brown, Joseph, Gordon McCord, and Paul Zachary. 2017. "Sunday, Bloody Sunday: Evidence from Northern Ireland for the Effect of Ethnic Diversity on Violence." Working paper.

Chen, Ted Hsuan Yun, Paul Zachary, and Christopher J. Fariss. 2017. "Who Protests? Using Social Media Data to Estimate How Social Context Affects Political Behavior." Working Paper.

Crowder, Kyle. 2000. "The racial context of white mobility: An individual-level assessment of the white flight hypothesis." *Social Science Research* 29 (2): 223–257.

Enos, Ryan D. 2014. "Causal effect of intergroup contact on exclusionary attitudes." *Proceedings of the National Academy of Sciences* 111 (10): 3699–3704.

Enos, Ryan D. 2016. "What the demolition of public housing teaches us about the impact of racial threat on political behavior." *American Journal of Political Science* 60 (1): 123–142.

Enos, Ryan D, and Noam Gidron. 2016. "Intergroup behavioral strategies as contextually determined: Experimental evidence from Israel." *The Journal of Politics* 78 (3): 851–867.

Fariss, Christopher J., and Zachary M. Jones. 2017. "Enhancing Validity in Observational Settings When Replication is Not Possible." *Political Science Research and Methods* https://doi.org/10.1017/psrm.2017.5.

Gee, Laura K., Jason J. Jones, Christopher J. Fariss, Moira Burke, and James H. Fowler. 2017. "The Paradox of Weak Ties in 55 Countries." *Journal of Economic Behavior and Organization* 133 (Janurary): 362–372.

Gelman, Andrew, and Donald B Rubin. 1992. "Inference from iterative simulation using multiple sequences." *Statistical science* pp. 457–472.

Gentry, Jeff. 2015. *twitteR: R Based Twitter Client*. R package version 1.1.9.
URL: *http://CRAN.R-project.org/package=twitteR*

Getmansky, Anna, and Thomas Zeitzoff. 2014. "Terrorism and voting: The effect of rocket threat on voting in Israeli elections." *American Political Science Review* 108 (3): 588–604.

Grace, Rob, Jess Kropczynski, Scott Pezanowski, Shane Halse, Prasanna Umar, and Andrea Tapia. 2017. Social Triangulation: A new method to identify local citizens using social media and their local information curation behaviors. In *Proceedings of the 14th ISCRAM Conference*. pp. 902–915.

Hasan, Samiul, Xianyuan Zhan, and Satish V Ukkusuri. 2013. Understanding urban human activity and mobility patterns using large-scale location-based data from online social media. In *Proceedings of the 2nd ACM SIGKDD international workshop on urban computing*. ACM p. 6.

Hauge, Michelle V, Mark D Stevenson, D Kim Rossmo, and Steven C Le Comber. 2016. "Tagging Banksy: Using geographic profiling to investigate a modern art mystery." *Journal of Spatial Science* 61 (1): 185–190.

Jiang, Bin, and Jean-Claude Thill. 2015. "Volunteered Geographic Information: Towards the establishment of a new paradigm." *Computers, Environments and Urban Systems* 53: 1–3.

King, Gary. 2013. *A solution to the ecological inference problem: Reconstructing individual behavior from aggregate data*. Princeton, NJ: Princeton University Press.

Kruse, Kevin. 2005. *White Flight: Atlanta and the Making of Modern Conservatism*. Princeton, New Jersey: Princeton University Press.

Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas A. Christakis, Noshir Contractor, James H. Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, and Marshall Van Alstyne. 2009. "Computational Social Science." *Science* 323: 721–723.

Lazer, David, and Jason Radford. 2017. "Data ex Machina: Introduction to Big Data." *Annual Review of Sociology* .

Neal, Radford M. 2000. "Markov chain sampling methods for Dirichlet process mixture models." *Journal of computational and graphical statistics* 9 (2): 249–265.

O'Leary, Mike. 2010. Implementing a Bayesian approach to criminal geographic profiling. In *COM. Geo*.

Rossmo, D Kim. 1999. *Geographic profiling*. CRC press.

Sands, Melissa L. 2017. "Exposure to inequality affects support for redistribution." *Proceedings of the National Academy of Sciences* p. 201615010.

Shadish, William R. 2010. "Campbell and Rubin: A Primer and Comparison of Their Approaches to Causal Inference in Field Settings." *Psychological Methods* 12 (1): 3–17.

Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2001. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Wadsworth Publishing.

Steinert-Threlkeld, Zachary C. 2017. "Spontaneous Collective Action: Peripheral Mobilization During the Arab Spring." *American Political Science Review* .

Steinert-Threlkeld, Zachary C., Delia Mocanu, Alessandro Vespignani, and James H. Fowler. 2015. "Online Social Networks and Offline Protest." *EPJ Data Science* 4 (19): 1–9.

Stevenson, M.D, Verity, and R. 2014. *Rgeoprofile : Geographic Profiling in R.* London, England: Queen Mary University of London. Version 1.2.
URL: *http://evolve.sbcs.qmul.ac.uk/lecomber/sample-page/geographic-profiling/*

Verity, Robert, Mark D Stevenson, D Kim Rossmo, Richard A Nichols, and Steven C Le Comber. 2014. "Spatial targeting of infectious disease control: identifying multiple, unknown sources." *Methods in Ecology and Evolution* 5 (7): 647–655.