

# THE EFFECT OF *NEW YORK TIMES* EVENT CODING TECHNIQUES ON SOCIAL MOVEMENT ANALYSES OF PROTEST DATA

Erik W. Johnson, Jonathan P. Schreiner  
and Jon Agnone

## ABSTRACT

*We know a great deal about the ways in which routines of news coverage may bias newspaper content, but little about how different article retrieval practices influence newspaper data assembled by scholars. Using the New York Times as a source of data on social movement activity, we compare depictions of protest by the African-American Civil Rights movement over time produced using the two most common article retrieval methods: index versus full-story coding. Full-story coding clearly offers more depth and greater breadth in terms of the events identified. Moreover, many of the same event characteristics associated with selection bias in newspaper reporting (e.g., size and confrontational nature of a protest event, presence of counter-demonstrators or police, and event sponsorship by a recognized social*

---

Narratives of Identity in Social Movements, Conflicts and Change  
Research in Social Movements, Conflicts and Change, Volume 40, 263–291  
Copyright © 2016 by Emerald Group Publishing Limited  
All rights of reproduction in any form reserved  
ISSN: 0163-786X/doi:10.1108/S0163-786X20160000040020

*movement organization) are selected upon again when stories are indexed by New York Times staff.*

**Keywords:** Social movements; civil rights; protests; newspaper; methodology; *New York Times*

## INTRODUCTION

Social movement scholars rely heavily on newspapers as a source of data for social movement activity – particularly for determining the occurrence and characteristics of protest events. The methodological implications of relying on newspapers have been both strongly critiqued (Myers & Caniglia, 2004; Ortiz, Myers, Walls, & Diaz, 2005) and defended (Earl, Martin, McCarthy, & Soule, 2004; Jenkins & Perrow, 1977). Such inquiries have focused primarily on selection and description biases that result from newspaper reporting routines (e.g., reporters mostly work regular work weeks) and event characteristics (e.g., size, whether the event is violent in nature or if police are present) (Earl et al., 2004; Olzak, 1989). This paper focuses on a specific type of understudied selection bias related to newspaper indexing.

As the national paper of record, the *New York Times* (*NYT*) is the most commonly used data source to assemble information on national social movements in the United States. Due to its long history, the *NYT* is an especially valuable source of information for longitudinal and historical analysis. Typically scholars do not read the entirety of the paper when identifying relevant social movement protest events. Instead, data are predominately drawn from the *New York Times Index* (*NYT Index*), consisting of short summaries of each article published in the paper, usually a sentence or two.

A large body of research has examined the characteristics of events, including protest events, which get selected for coverage in media outlets. Indexing bias, a specific type of selection bias introduced when news stories are reduced to a few condensed sentences summarizing a more elaborated story, is less studied. Indexing bias has primarily been of concern to political scientists interested in agenda setting within media (Baumgartner & Jones, 2002; Woolley, 2000). Only a very small body of research, primarily in international comparative literature, has explored indexing bias in event

counts (Taylor & Jodice, 1983). We are aware of no empirical test of indexing bias on the key measure of social movement activity: collective action events. We assess the implications of indexing bias in the study of social movement protests by posing and attempting to answer two research questions. First, do social movement protest event *counts* appreciably differ when assessed through index coding versus full-story coding? Second, how do the *types* of protest events captured in full-story versus index-based coding differ?

To answer these questions, we compare protest data counts from the *NYT* collected via *index coding*, as utilized by McAdam (1982), Olzak (1989), and Jenkins and his students (Agnone, 2007; Jenkins, Jacobs, & Agnone, 2003; Jenkins & Perrow, 1977), with *full-story coding* from *The Dynamics of Collective Action* (DCA) research project conducted by Doug McAdam, John McCarthy, Susan Olzak, and Sarah Soule. Publications in major sociology journals have drawn extensively from the publicly available DCA dataset, and the rich details on event characteristics contained therein (Beyerlein, Soule, & Martin, 2015; Davenport, Soule, & Armstrong, 2011; Earl & Soule, 2010; Earl, Soule & McCarthy, 2003; Johnson, 2008; Johnson, Agnone, & McCarthy, 2010; Jung, King, & Soule, 2014; King, Bentele, & Soule, 2007; Martin, McCarthy, & McPhail, 2009; McAdam & Su, 2002; Olzak & Soule, 2009; Soule & King, 2008; Walker, Martin, & McCarthy, 2008; Wang & Soule, 2012). By leveraging these existing datasets we are able to undertake the first empirical assessment of indexing bias in the *NYT*. Our case study is the African-American Civil Rights movement, an ideal case given the extensive extant literature and its central influence on the development of social movement theory over the past 40 years (Jenkins & Eckert, 1986; McAdam, 1982; Meyer & Minkoff, 2004).

After reviewing literature on newspaper protest event data, we compare the results of index and full-story coding in the yearly number of African-American Civil Rights protest events from 1960 to 1995. The two data access methods produce very different overall counts of Civil Rights protests, with the full-story coding identifying more than three times as many events as index coding. While highly correlated early in the observation period, when rates of protest activity are exceptionally high, there is only a weak association between the datasets late in the period when protest activity is low. A comparison of a subset of articles that make it into the full-story and index datasets shows that many of the elements of a protest that make it more likely to be covered in a newspaper to begin with are selected upon again at the indexing stage. Events that are larger, sponsored by an SMO, adopt protest tactics which are more disruptive, or have dramatic

conflict between parties (e.g., presence of police or counterdemonstrators) are particularly likely to be identified when relying on index coding versus full-story coding. Thus, between the reality of events and index-generated counts there exists a *double filter*. The paper concludes with a discussion of the implications of this double filtration process for social movement research and a call for further research into how both of the newspaper access methods we look into here compare to the newest research strategy of full text (often automated) keyword searches.

### **SOCIAL MOVEMENT EVENT COUNTS, NEWSPAPERS, AND THE *NEW YORK TIMES***

Newspaper-based event data have long been a workhorse for social movements research examining collective action dynamics over time and have played an important role in the development of political opportunity theories about movement emergence (Jenkins & Perrow, 1977; McAdam, 1982), studies of changing tactical repertoires and their diffusion (McAdam, 1983; Olzak & Uhrig, 2001), protest policing (Davenport et al., 2011; Earl et al., 2003), and of social movement outcomes (Burstein, 1985; Johnson et al., 2010; King et al., 2007; McAdam & Su, 2002). Here, we briefly summarize what researchers already know about biases in newspaper data. We then introduce how different methods of accessing newspaper data (index vs. full-story) may affect information collected.

#### *Biases in Newspaper Data*

Critics of using newspaper data in social movement empirical analysis have focused on the existence of description and, especially, selection biases in newspaper coverage (Myers & Caniglia, 2004; Ortiz et al., 2005). Assessments of newspaper description bias, or the accuracy of newspaper reports, have generally found a high degree of correspondence between news coverage and reports by independent sources for “hard” news items (e.g., the form of an event) when they are reported (Earl et al., 2004; McCarthy, McPhail, & Smith, 1996). More problematic is descriptive bias for “soft” news items (i.e., impressions and interpretation of events by reporters). Perhaps in part as a response to this, movement scholars have

tended to focus attention on “hard” news items, the form and goals (i.e., claim) of an event.

Selection bias refers to potential bias in the subset of events actually reported in newspapers, and is potentially more problematic in comparing full-story versus index coding. Clearly, not all social movement activities are created equal in terms of the likelihood that a newspaper will cover them, with some types of events (especially large, contentious, and spatially proximate ones) more likely to be defined as “newsworthy” and thus included in final print editions than others (Earl et al., 2004; McCarthy et al., 1996; Oliver & Myers, 1999). The alignment between news outlets and protestor claims is also relevant. There are, for example, considerable differences between the *NYT* and ethnic newspapers in the coverage of civil rights protests (Weiner, 2011). Our analyses do not speak to these well-studied aspects of bias in media generally, or the *NYT* in particular.

#### *Methods of Social Movement Event Data Collection*

Instead, we look at the effect of differing approaches to accessing *NYT* data, and implications for conclusions about the pace and dominant forms of protest activity. The most common method of accessing the *NYT* is to scan brief summaries of published articles via the *NYT Index*. These summaries, as short as a sentence and typically no more than two or three sentences, are written by long-serving staff that are trained and supervised in an effort to maintain high quality and consistency over time. Index abstracts are written for nearly all *NYT* stories, excluding society news and “question and answer” columns, as well as most letters to the editor prior to 1997 (see Althaus, Edy, & Phalen, 2001, pp. 710–711 for details on what is systematically and intentionally excluded from coverage in the *NYT Index*).

Over the past several decades scholars have increasingly moved toward full-story coding of the *NYT* and other news sources, yet indexes remain a prominent information source. For researchers, the benefits of coding the activity of a particular social movement from the *NYT Index* (or another media index), as opposed to full-story, can be considerable. The index approach is a less resource intensive and more straight-forward task. There is less information to process with an index entry than in the full-story format, given that an independent third party has already tagged and grouped articles according to relevant topical keywords. Rather than reading entire newspapers to identify events of interests, researchers can focus on

identifying appropriate index headings (e.g., Civil Rights Movement, Black Power Movement, and so forth) and then read only the relevant article summaries to identify events of interest.

Until fairly recently scholars relied almost exclusively on newspaper indexes. In particular, as noted by Olzak (1989, p. 130), the early and influential studies of “race riots and civil rights’ activity used only an annual index, usually from the *NYT*” (Burstein, 1985; Jenkins & Eckert, 1986; McAdam, 1982; Spilerman, 1970). More recent civil rights research has continued to draw on *NYT Index* data (Jenkins et al., 2003; Meyer & Minkoff, 2004; Weiner, 2011) even as full-story search methods grow increasingly popular.

It has been political scientists (Althaus et al., 2001; Woolley, 2000) who have most persuasively elucidated the inherent limitations of relying on indexes, including the *NYT*, as opposed to the full stories of daily newspapers. This owes, at least in part, to the likely greater effect of indexing bias for media agenda setting research than for event counts (Woolley, 2000, p. 162). Three characteristics of index listings stand out. First, an index is by design considerably briefer than the complete article, containing only the title and maybe a one or two sentence summary of an article. This leads to the expectation that searches of index material are likely to produce lower event counts than searches in entire *New York Times* daily editions. In our review of the literature, we found only one test of this event inclusivity hypothesis. Taylor and Jodice (1983, p. 184) find that 29% of a broad spectrum of international political events (including demonstrations and riots, as well as armed attacks, assassinations, political strikes, and government sanctions) found in daily editions of the *NYT* are identified through index coding. Armed attacks and political assassinations are particularly likely to be identified through index coding, while demonstrations and riots are under-reported in the index relative to full story coding.

A second concern with indexing is the consistency of index terms over time. Index stories are listed in keyword categories that act as selection schemas which filter article retrieval, but relevant keywords may not even exist, or may change over time. Baumgartner and Jones (1993), for example, in their study of public agenda setting found the entry “child abuse” appeared in the *Readers Guide to Periodical Literature Index* beginning in 1959. This could be construed as indicating minimal attention to the issue previous to this date. Using the *NYT Index* as a second source, however, Baumgartner and Jones were able to show substantial attention to the problem of child abuse in the 1920s as well. This finding speaks to the high quality of *NYT Indexing*. It also suggests that for emerging social

movements (or public issues) still early in the mobilization phase, that designated keywords of relevance are less likely to exist in an index and thus for relevant events to be located.

Moreover, where relevant categories exist, they may shift over time. Large and influential social movements, like civil rights, produce widespread cultural change that is reflected in the language (Eyerman & Jamison, 1991). What was once the “black movement” became “civil rights – black” and later “African-American civil rights.” Althaus et al. (2001) describe a seismic shift in the *NYT Index* that occurred in 1983 and had more to do with the routines of work than cultural changes. In 1983, the thesaurus containing the comprehensive list of subject terms from which indexers work went from approximately 25,000 subject terms to 3,500. Little is known about how the emergent nature of index subject coding translates to the analysis of protest.

Finally, the intent of an index is to reduce information. Index recordings of events provide researchers with less information than if they were working with full-stories. This suggests that a focus on aggregate protest is likely to be less problematic than attempts to use index’s to parse protests into more specific categories based on, for example, the dominant protest form or the target of protest (Walker et al., 2008). In a similar vein Althaus et al. (2001), in a comparison of policy statements derived from all index headings under “Libya” during the 1986 Libya crisis to the same set of full stories, find that find policy statements in favor of the administration are similar at an aggregate level across the two sources. The validity of indexing breaks down, however, when statements are disaggregated by source (e.g., foreign official, US Congress, and US Administration).

As the limitations of indexes have become more apparent, some researchers have moved toward full-story article coding to identify relevant events of interest. This transition has been accelerated by technological advances and the 2010 public data release by the Dynamics of Collective Action (DCA) project, funded by the National Science Foundation.<sup>1</sup> The shift toward full-story access builds on the assumption that full-story it is a more comprehensive source than index searches. Rather than take this as a given, we consider it an open question and therefore provide the first empirical assessment therein. What are the implications for the large body of outstanding work that relied heavily on indexes, and future work which continues to draw upon it? Are full-story results comparable with work relying on indexes to generate information? We leverage these new data from DCA to address these questions for social movement events specifically.

## DATA

In this paper, we are interested in comparing the two established primary methods of enumerating social movement activity via the *NYT* to determine how methodological decisions about data access may impact upon our understanding of collective action. As such, we conduct no new content analysis but instead rely on secondary analysis of datasets from two independent, established projects that feature prominently in the social movement literature. Each of these projects assembled information from the *NYT* to enumerate African-American Civil Rights protests, with one project relying on keyword searches of the *NYT Index* and the other searching full stories in the print version of the *NYT*.

Information on African-American Civil Rights movement protests assembled from *The New York Times Index* (*New York Times*, 1960–1995) was generously made available for analysis by Craig Jenkins (Jenkins et al., 2003) who updated and extended time-series data originally assembled by McAdam (1983) for the 1955–1970 period. This project culminates nearly 25 years of work involving at least 7 different coders (for details on reliability and coding procedures see Jenkins et al., 2003). In this dataset, social movement events are categorized into one of six discrete types.<sup>2</sup> While all social movement activities are meaningful, of primary interest in this study, and in much of the social movement literature, are protest events. Protest is defined in these data as nonviolent direct action events about African-American issues, including public demonstrations and marches, sit-ins, rallies, freedom rides, boycotts, and other direct actions. All protest events are coded as discrete, meaning multiday protests are considered a single event unless reported as distinct events with different actors and initiation.

In the second dataset, The Dynamics of Collective Action Project (DCA) has assembled and made available social movement event information based on content analysis of all collective action events reported in full-text daily editions of the *NYT* from 1960 to 1995. To be included in this dataset an event had to be collective in nature, have articulated some claim (either a grievance against, or expression of support for, some target) and must have occurred in the public sphere or have been open to the public (for project descriptions see Earl et al., 2003; McAdam & Su, 2002; Van Dyke, Soule, & Taylor, 2004). Each event is assigned up to four issue claims codes to allow multi-purposed actions. So, for example, a protest of environmental injustice in an African-American community can be coded as both an environmental and a civil rights event. All events coded as



African-American civil rights for any of up to four claims were selected to insure complete coverage.<sup>3</sup>

For our analyses we first used the DCA “form” and “action” variables to construct a more theoretical meaningful set of 16 tactical indicators (Ring-Ramirez, Reynolds-Stenson, & Earl, 2014), then selected relevant tactics to generate an operational definition of non-violent protest events comparable to that employed by Jenkins et al. (2003). In all, nine of the tactical categories created from this DCA recoding are used to identify events that are comparable to those identified in the *Index* data, including the non-radical tactics of Rally/Demonstration, March, Vigil, Picket, Ceremony, Dramaturgical demonstration, Boycott, and the radical tactics of Civil Disobedience (including sit-ins) and Strike/Slow Down. While scholars may conceptualize protest in many different ways (e.g., considering boycotts as an example of institutional activity rather than a protest activity), these epistemological debates are less important for this study than matching operational definitions of protest across the *Index* and full-story projects, to allow for an examination of whether different data access methods produce comparable protest samples.

## COMPARING AGGREGATE COUNTS FROM FULL-STORY AND INDEX-BASED CODING

We begin our analysis of the *NYT Index* and full-story coding by comparing the aggregate protest counts each data collection method elicits (Fig. 1). In total, over the entire observation period (1960–1995), *NYT Index* coding identified 2680 fewer protest events ( $N = 1,239$ ) than did full-story coding ( $N = 3,919$ ). This works out to a “capture” rate of 31.6%, strikingly similar to the 29% capture rate found by Taylor and Jodice (1983, p. 184). More important than aggregate capture rates for the type of over-time analyses that protest data are typically leveraged for, is if the sources trend together.

The event lines plotted in Fig. 1 share the same overall trajectory of decline after the 1960s, but also show some distinct differences. Early in the observation period, there are different peak years of protest incidence identified by the two sources. There is also a modest revival of protest shown during the 1990s using full-text data that is not apparent in the index.

To further comparison of the protest trends that derive from these two databases we next assess correlation coefficients between our data sources

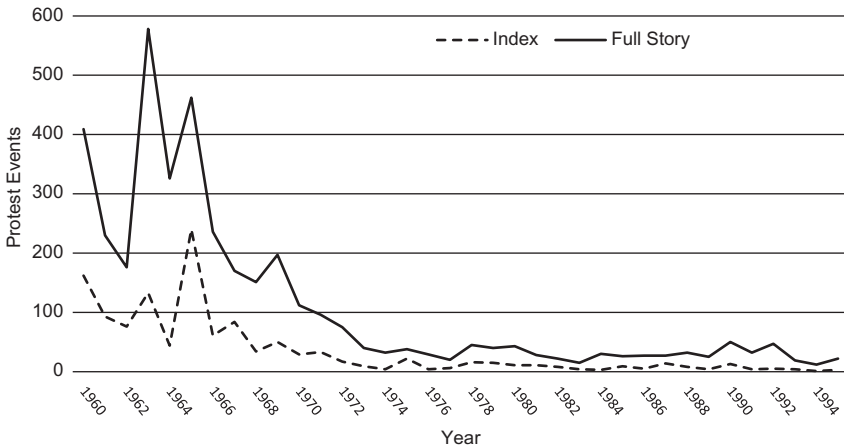


Fig. 1. Civil Rights Protest Events by the Year and Data Access Method, 1960–1995.

**Table 1.** DCA to *NYT* Index Correlations by Time Period.

All Years	12-Year Segments		
1960–1995	1960–1971	1972–1983	1984–1995
0.8964	0.7708	0.643	0.4945

overall, and across consecutive 12-year periods (Table 1). Over the entire observation period, protest counts generated by the index and full-story access methods are strongly correlated at .875. This strong overall association is driven by the steep decline in protest incidence after the 1960s in both sources, however. During the first third of the observation period, when the decline occurs, the correlation between data sources remains a robust .77. The correlation declines to .64 over the next 12 years, and further to a modest .49 from 1984 to 1995.

Plotting the ratio of full-story to index event counts (Fig. 2) provides a more dynamic picture of the relationship between the two data access methods. The ratio of full-story to index events generally hovers between about 2:1 and 4:1 from 1960 to 1983 (with notable exceptions in 1965, 1974, and 1976). Late in the observation period (7 out of the last 11 years) the ratio is higher than 5:1. The average ratio per year prior to 1984 was

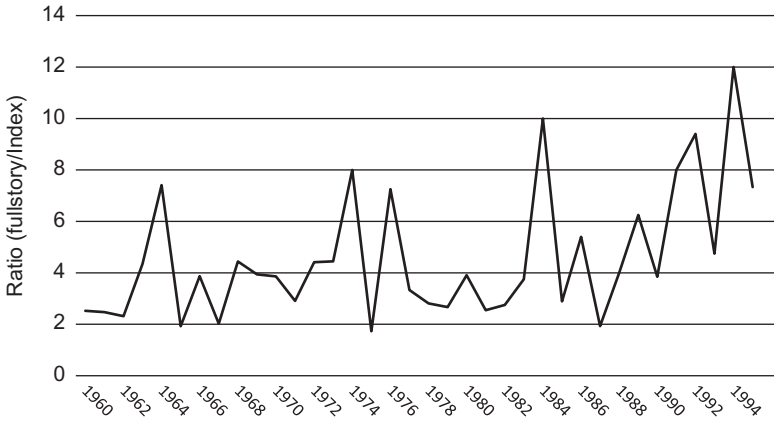


Fig. 2. Event Ratio (Full Story/Index) by Year.

3.73:1 while it is 6.32:1 from 1984 to 1995, again indicating weaker correspondence between the trend lines in the more recent, less mobilized, and period.

## COMPARING THE CHARACTERISTICS OF EVENTS IN FULL-STORY AND INDEX-BASED COUNTS

How do the characteristics of protest events contained in the *NYT Index* data compare with the fully story coding? To answer this question, we move beyond the comparison of aggregate yearly protest counts to analyze the content of relevant events.

The task of identifying, for matching purposes, exactly which *NYT Index* article summaries were coded as containing protest events proved exceedingly difficult for several reasons. First, Jenkins data were a continuation of McAdam's (1982) coding, with the McAdam series only available as aggregated spreadsheets minus the original coding materials. The Jenkins data team continued this aggregated data store with the intent to provide maximum comparability to the original McAdam series. Secondly, this means that Jenkins index coded data are generally available as counts of events by year, with only certain subsections of the data available with more detailed event coding (e.g., with event size and issue codes). Information on precisely which index stories are coded in a

particular way is not generally available, however, complicating the matching process considerably.

For a limited number of years Jenkins was able to locate and generously supplied for analysis original coding files that showed the dates events were coded. From this information, and the more detailed information contained in the DCA, we were able to carefully match across the two data sources for a four-year period (1977–1980). For this analysis, we are primarily interested in what aspects of a protest activity may trigger the event to be captured by the full-story coding strategy as compared to the index strategy. This matching process was performed not at the level of events, but on newspaper articles coded.

In both the Jenkins and DCA projects collective action events are the unit of analysis, not newspaper articles, meaning multiple events may be coded from a single article or index summary. Multiple events are more likely to be identified in a full-text reading of articles than an abstract summary. In the four-year period for which we are able to conduct analysis of the details of events located in the Jenkins and DCA data we find that that 17% of events listed in the DCA derive from the coding of multiple events from a single article (43 events across 17 articles). Only 12.3% of the events Jenkins identifies are the result of multiple events in an article (17 events across 8 articles). In most cases, these multiple events shared the same form (e.g., marches in two different cities). In cases where the DCA was able to tease out two or more different events with different forms (e.g., a march in New York City and a boycott in San Francisco, California) the article was coded as having aspects from multiple events (e.g., both a march and a boycott).

To help elucidate the differences between index and full-story coding, we provide two screenshots of a Saturday, July 21, 1979 protest event from Birmingham, Alabama that was coded in both the Jenkins and DCA data. We chose this event for several reasons. First, the event falls outside the period of heightened civil rights movement activity, thus we can observe how the media tracks events during more routine periods of political unrest. Second, we chose an event that is especially likely to be covered in the NYT Index. Our results suggest that the presence of police, large turnout, named SMO sponsorship, and dominant event form (a march) of this case are all conditions that enhance correspondence.

The full-story, highlighted below, took up nearly a third of the real-estate of page 6, replete with a large photograph of the event proceedings. The article, spanning 16 paragraphs, contained approximately 920 words and identified several organizations that were involved, including the

NAACP, the SCLC (Southern Christian Leadership Coalition), and the Ku Klux Klan. The story explained in detail the background grievances related to police brutality within the Birmingham community of color, most recently against a 20-year-old African-American woman who was killed by police, includes interviews with a few protest participants and government officials, and provided historical context on this being the largest protest in the city since the 1963 demonstration led by Dr. Martin Luther King.<sup>4</sup>

In comparison to the *NYT* full-story article is the *NYT Index* entry which reads as follows: “some 2,000 blacks march peacefully through downtown Birmingham in protest (M), J1 21,6:3” (*NYT Index*, 1960–1995, p. 159). This entry, indexed under the keyword heading *Blacks (in the United States)*, is couched within the larger bucket of related topical issues that occurred in the state of Alabama (some issues of the *NYT Index* split events by state when there exists a multitude of topics under the keyword heading). The available information in the index is, by design, much more terse than the full-story, as this example showcases. It still manages, however, to convey much of the important information conveyed in the full story, including the size, timing, and location of the event, as well as to accurately describe it as a peaceful protest by “blacks” around the issue of police brutality. The connection to police brutality, however, is only made possible through a very careful reading of the *NYT Index* by a trained observer willing to impute a connection between this and the previous entry which summarizes a July 18 article that notes plans for a March in Birmingham to protest an “unjustified” shooting by a white police officer. The *NYT Index* does not report the participation of three social movement organizations (including the NAACP, SCLC, and Black Muslims) or account for the presence of police (and the absence of arrests), all potentially important information that the DCA full-story coding is able to glean.

In Table 2, full-story and index identified events are arranged categorically by dominant protest tactic. Of all the articles captured in the full-story coding, 31.4% are also identified in the *Index* assembled data. Using a two-tailed *z*-test to compare proportions, marches are significantly more likely to be picked up in index coding, with half of the events that adopted the form of a march also identified through *Index* coding. On the other hand, acts of civil disobedience which tend to be smaller in scale are perhaps less likely (significant at the .1 level) to be picked up in Index-based coding. Only 4 out of the 27 civil disobedience events from the full-story project were coded as such in the Index.

**Table 2.** DCA Full-Story and Matched *NYT Index* Events, by form, 1977–1980.

Event Type	Full-Story	Index	Percent Full-Story in Index
Total event forms	188	59	31.4
Rallies	55	14	25.5
Marches	48	24	50**
Ceremonies	2	1	50.0
Vigils	3	2	66.7
Pickets	15	6	40.0
Civil disobedience	27	4	14.8*
Strikes	19	4	21.1
Boycotts	19	4	21.1

\* $p < .1$ , \*\* $p < .05$ , \*\*\* $p < .01$  (two-tailed  $z$ -test).

What Table 2 does not make apparent is the ways in which racially based strike and boycott activity are linked, and differentially identified when drawing primarily from the *NY Times Index* instead of full-story coding. In total, 21.1% of all *New York Times* news stories reporting on black civil rights motivated strike and boycott events are also identified when relying on the index, not a significant difference. However, boycotts and strike activity, at least in this issue domain, are almost never reported in isolation but instead almost always occur as part of multi-form events. One hundred percent of strike events in our sample, and all but one boycott event, co-occur. Strikes and boycotts are reported as parts of the same event the vast majority of the time (16 out of 19 total strike events co-occur with boycotts, and vice versa). In half of those cases (8) newspaper stories report strikes and boycott activity as co-occurring in isolation. In a quarter of cases, four events, strikes, and boycotts co-occur along with marches. These four cases are the 21.1% of strikes and boycotts identified when relying on the index. In other words, index coding picking up only those articles where strikes and boycotts co-occurred both with one another AND a march. As the only protest forms with direct economic consequences, boycotts and strikes challenge dominant power structures and are less likely to be covered by newspapers in the first place (Herman & Chomsky, 1988). When they are covered, our evidence suggests that they are buried in stories that feature complementary protest forms with a high degree of newsworthiness (i.e., marches).

**Table 3.** DCA Full-Story Counts and Matched *NYT Index* Events, by Characteristics, 1977–1980.

Event Characteristic	Full-Story	Index	Percent Full-Story in Index
Total entries	124	35	28.2
Front page	9	3	33.3
New York City location	21	7	33.3
Size of crowd mentioned	77	27	35.1
Crowd 1,000 or more	13	6	46.2*
Crowd 100–999	40	15	37.5
Crowd under 100	24	6	25.0
No crowd size mentioned	47	8	17.0*
Police mentioned	41	20	48.8***
No police mentioned	83	15	18.1**
Counter demonstrators	15	8	53.3**
No counter demonstrators	109	27	24.8
Identified SMO	49	19	38.8
No SMO identified	75	16	21.3

\* $p < .1$ , \*\* $p < .05$ , \*\*\* $p < .01$  (one-tailed  $z$ -test).

Other event characteristics matter for understanding over/under representation in *Index* relative to full-story coding. This can be seen in Table 3. One-tailed  $z$ -tests are used here because we have theoretical reasons to expect directions of effect, based on previous research on event selection bias in newspaper coverage generally. Events where police are present, that are larger, and where counterdemonstrators are present are all particularly likely to be identified when relying on the index, as are events sponsored by an identified social movement organization. Index data may roughly proxy the relative incidence of protest activity in times of heightened mobilization, but the depiction of what protest looks like in the *Index* relative to full-story *NY Times* is more violent, more contentious and larger.

## RESULTS – REGRESSION

Finally, we test how these differences play out in multivariate regression models of protest frequency. We draw heavily from Jenkins et al. (2003)

for guidance in identifying theoretically important covariates, including: the African-American unemployment rate, the ratio of black/white household income, number of Vietnam War casualties, and annual membership of the NAACP (National Association for the Advancement of Colored People). Elite political allies are measured as the sum of congressional seats held by African-Americans, the total number of non-southern Democrats in Congress, and a dummy variable coded “1” during election years with an incumbent Republican president. Divided government is operationalized as a dummy variable representing the presence of divided party control over the Senate, the House of Representatives, and the Presidency (coded 1 when any of the Senate, House of Representatives or Presidency are in majority control of different parties; 0 when all three are majority controlled by the same party).

In addition, we control for several factors not in [Jenkins et al. \(2003\)](#). Data on Congressional Hearings from the Policy Agendas Project (PAP) Congressional Hearings database are used to assess the pace of federal policy debate.<sup>5</sup> We reviewed the PAP provided summaries of each hearing in relevant topic areas to identify a total of 435 congressional hearings on African-American civil rights during the time period.<sup>6</sup> Public opinion is measured using [Stimson’s \(\[1991\] 1999\)](#) measure of policy mood, a time series measure of public support for government programs on the liberal-conservative continuum.

[Table 4](#) presents results of models examining whether differences in protest event counts between the two sources of Civil Rights movement data impact the results of theoretically motivated regression models. Poisson regression models are appropriate statistical inference models when count data are used as the dependent variable ([Long, 1997](#)), and common within the relevant literature ([Agnone, 2007](#); [Meyer & Minkoff, 2004](#)). Still, yearly count data of a limited number of observation points (36 years) may readily result in over-specified regression models. Given this limitation, and that we are interested in the parallels or differences between models, rather than the substantive topic of African-American protests per se, we focus on a summary of the overall patterns in results. Interested readers will find descriptive statistics, bivariate correlations, and full-regression models with coefficients in the [appendix \(Tables A1–A3\)](#).

In our models using protest counts derived from the DCA full-story coding and the *NYT index*, the majority of independent variables are comparable in effect regardless of data collection strategy. However, measures of public opinion, black unemployment, congressional attention to civil rights issues, and NAACP membership, each of which account for theoretically



**Table 4.** Summary Results from Models Regressions Various Independent Variables on Protest Counts Derived from Index and Full-Story Coding Strategies.

Variables	Difference between Index and Full-Story	Index	Full-story
Black/White income gap	No difference – consistent	–*	–*
Northern democrats in congress	No difference – consistent	+*	+*
Divided congress	No difference – consistent	+*	+*
Black congressmen	No difference – consistent	–*	–*
Vietnam war deaths (LN)	No difference – consistent	+*	+*
Protests ( $t-1$ )	No difference – consistent	–*	–*
Republican incumbent president	Changed Direction AND significance	–*	+
NAACP membership (LN)	Changed Direction AND significance	–	+*
Black unemployment	Changed significance	–	–*
Public opinion ( $t-1$ )	Changed significance	+	+*
Civil rights hearings ( $t-1$ )	Changed significance	+	+*

\*Significance at the  $p < .05$  level.

important concepts in an analysis of social movement dynamics and which are significant in the expected direction in full-story models, fail to achieve statistical significance as predictors of protest activity in index-based models. These measures have only a very small amount of model explanatory power, when comparing the small changes in Pseudo  $R^2$  across the nested models, but the changing levels of significance for them (as well as Republican president), suggest that different data access strategies may alter conclusions about the fundamental drivers of social protest.

## DISCUSSION AND CONCLUSION

Our findings paint a clear picture: that full-story coding of the NYT results in the identification of a greater number of events than does coding entries in the *NYT Index*. Moreover, indexing results in the application of a double filter. The newsworthiness of social movement protest is filtered first at the newspaper reporting stage, when reporters and editors disproportionately write about events that are, all other things equal, large, disruptive, or imbued with some sort of conflict. This resulting sample of reported protest events is, once again, filtered on many of the same criteria by newspaper

staff in the course of indexing. The application of systematic filters during the indexing process results in the types of disruptive and conflictual events that Tarrow (1989) claims characterize high-activity periods of protest cycles particularly likely to be identified. Protest events that occur outside periods of heightened protest are typically more routinized, and less likely to be picked up by the *NYT Index*.

Newspapers remain an important source of data for social movement researchers, but the way they are accessed is changing rapidly. It was only a few years ago that Doug McAdam, John McCarthy, Suzan Olzak, and Sarah Soule first made hand coded data on protest activity from full-story searches of the *NYT* widely available through the Dynamics of Collective Action (DCA) project. Prior to that, the vast majority of analyses relied heavily on index searches.<sup>7</sup> This begs the question: How much scholarly work has relied on these various data access methods whether scholars have tended to utilize one more than the other over the past decade? To help answer this question we put together, to the best of our ability, a list of published social science scholarship utilizing the *NYT Index* or DCA to establish collective action data series (see online supplement at the lead authors website or contact the authors for details). Table 5 summarizes our findings. Overall, we identified 64 articles using *NYT Index* data (1977–2014) and 36 articles relying on DCA data (2002–2015), spanning publication across the top journals in sociology and beyond. Publications since 2005 include 26 of the 63 (41.3%) of the articles using *NYT Index* data, and 29 of the 32 (90.6%) of the articles using DCA data. This suggests that the DCA and other full-story access methods are increasingly popular, but that both data access methodologies continue to be relied upon.

The increased digitization of a broad range of materials, and the proliferation of methods for electronically coding those materials, practically assures increased scholarly usage of full-story records of newspapers and

**Table 5.** NYT Index and DCA Published Articles, Book Chapters, and Books.

Data Source	Published Articles and Book Chapters/Books	Earliest Publication Date	Latest Publication Date	Published Since 2005
NYT index	63	1977	2014	26 (41.3%)
DCA	32	2002	2015	29 (90.6%)

other documents in the future, but with some implications. For example, in one of the seminal pieces considering differences between print and digital newspaper access, Nicholson (2013, p. 64) notes that “a hard copy of a newspaper is fundamentally different from a digitized version,” pointing out that digital papers need to be studied more in terms of how they can be accessed and searched in comparison to traditional print newspapers. Electronic databases often contain stories that vary from the print source (Ridout, Fowler, & Searles, 2012). The Reference Specialists in the Newspaper & Current Periodical Reading Room at the Library of Congress (personal communication) points out that, “There is some variation between the online editions of newspapers, including the *New York Times*, and print editions (and, of course some variation within various print editions, as well). For example, we have found that at least a few online articles do not appear in print editions. Also images, titles, and text may vary, even when articles appear in both.”<sup>8</sup> The DCA full-story coding strategy produced over three times more protest events than were found in the *NYT Index*. These two methods of data access were highly correlated during periods of robust protest activity, and less so during periods of relative movement quiescence. This is, in part, a product of the fact that the types of events that are more likely to be noted as protests in coding the *NYT Index* – large events and marches – were more prevalent during the heightened mobilization phase of the Civil Rights movement.

While there is a clear double-filtering process that occurs as a result of indexing, the implications for analyses of social movements during periods of heightened mobilization and conflict, and high levels of aggregation, are minimal. In regression models predicting yearly protest counts the majority of independent variables were comparable in effect across our models, regardless of data collection strategy. This is reassuring and suggests that, at least at high levels of aggregation, researchers can reasonably rely upon either full-story or index-generated data on movement protest. Given the ease of accessibility and relative speed of data collection, index coding should remain a reasonable research accommodation during periods of heightened movement activity and/or when resultant data are part of a study in a less central way; for instance, as an independent control rather than as the dependent measure of central analysis as here.

Models did exhibit some important differences, however, and had a better fit to the data when using full-story as compared to index-generated protest counts as the dependent variable. This reinforces the notion that full-story data access and coding is generally preferable over index-generated data when possible and practical. While index coding may act as

a valid proxy for protest activity captured in full-story articles, the full-story coding strategy appears to provide a more complete depiction of protest activity, which is relevant to researchers engaged in comparing competing theories on their relative merits.

Protest has long been the most studied indicator of social movement activity, and the *New York Times* remains an important source of information when generating protest counts. With the proliferation of full-story search capabilities researchers may choose to rely on index searches with increasing rarity, instead opting to take advantage of the newer and more complete full-story datasets. This has advantages, but also creates difficulties in terms of accumulating knowledge. Scholarship produced from full-story content coding will inevitably be compared to scholarship analyzing index-coded data. We have shown here that different conclusions may be drawn depending upon choice of data access strategy. New data access capabilities, including full-story coded data, can help researchers understand the social world, but researchers must be careful when drawing comparisons to research based on different data as findings may be a product of an index bias rather than something more substantively meaningful.

Our results also suggest another potential difficulty for researchers attempting to assemble time series data, as is common in the study of social change. As full-story materials become increasingly accessible in the digital age, it is likely that researchers will encounter situations where full-story access for some source materials are available for the recent period, but only index data in the earlier period (a situation not unlike what we encountered in writing this paper: DCA data are available only after 1961, but Jenkins provided index coding much earlier). Our research suggests that data collected with index-coded procedures is different enough from data collected with full-story coded procedures to complicate comparisons and undermine longitudinal research. Both types of data collection may be valid, but extreme caution should be used in any attempts to incorporate them together in an effort to assemble a longer time-series.

There are, as always, a number of limitations to the study results presented here. A primary concern is the issue of generalizability. We examine only one social movement and one type of event (protest). While we expect the results produced here to be generalizable beyond the African-American Civil Rights movement, the extent to which they are remains an open empirical question.<sup>9</sup> Moreover, there is a vast literature that uses newspaper article counts as a proxy for media attention to an issue. The high correlation between index and full-story coding of protest events provides some confidence, but additional research is needed to assess how accurately

index counts trace media attention to issues more generally (though see Woolley, 2000).

Additionally, movement scholars have given increased attention in recent years to a broad range of routine institutional events (e.g., filing lawsuits, petitioning, holding press-conferences, and other information distribution activities) that do not meet traditional definitions of protest activity. Political mediation theory, for example, argues that the fit between the political environment and the *type* of activities movement adopt (i.e., disruptive protest or institutional influence activities) is a major factor in explaining movement success (Andrews, 2001; Amenta, Halfmann, & Young, 1999; Johnson et al., 2010). Research on newspaper selection bias (see Earl et al., 2004 for a summary) has consistently shown newspapers to be more likely to report on events that are seen as relatively more “newsworthy.” Factors that enhance the newsworthiness of events (including the size and intensity of an event, and the presence of counterdemonstrators or police) generally correspond with the characteristics of more disruptive protest. The more routine institutionalized activities which social movements regularly participate in, and which are an increasing focus of concern for social movement scholars, are particularly likely to be underrepresented in index, versus, full-story coding of newspapers.

There is a need for greater attention to the alternative ways in which newspaper data are accessed and processed, and how that may influence the comparison of scholarly results. There are credible statistical corrections that compensate for undercounting and selection biases in numerous sources of social science data, most prominently for the U.S. Census (US Bureau of the Census, 1980), but no such statistical correction strategies exist for selection bias in newspapers (indexed or not). We are not alone in identifying the need to develop such procedures (Jenkins & Maher, 2016) and hope that this work contributes to attempts to control for non-randomness in the selection of newspaper event data.

Moreover, while the DCA data have been very influential in the social movement field in a short time, they are already proving themselves to be anachronistic in some ways. Computing technologies were important in their assembly, but in the end the coding was done by trained undergraduate research assistant reading each newspaper and content coding information into electronic databases. Shortly after publication of these data, the *NYT* made available an alternative method of obtaining full-story over time search results built around the use of keyword searches. The ongoing revolution in digital technologies and the processing of text-based data in particular, represented, for example, in the growing adoption of automated

text classification procedures (Hopkins & King, 2010), suggests a need for considerable research into additional biases that may result from the coding of digital versus print versions of the *NYT*.

## NOTES

1. Data can be found at the Dynamics of Collective Action (DCA) website: [www.stanford.edu/group/collectiveaction](http://www.stanford.edu/group/collectiveaction)

2. Jenkins et al. (2003) coded for six categories of events: (1) Organizational Changes (such as the formation or merger of a group); (2) Information events (such as a press release or a speech); (3) Resource Mobilization Drives (activity centered around the raising of money or planning events); (4) Institutional Political Action (includes directly engaging the government through lobbying efforts or proposing legislation); (5) Protest, defined as nonviolent direct action events by African-Americans, including public demonstrations and marches, sit-ins, rallies, freedom rides, boycotts, and other direct actions; and, (6) Violence, which included events such as riots, melees, and racial confrontations that lacked a clear protest quality.

3. A complete codebook, including a full list of issue claim codes, can be found on the Dynamics of Collective Action (DCA) website: <http://web.stanford.edu/group/collectiveaction/cgi-bin/drupal/>

4. A digital reproduction of the original page layout is available via the *New York Times* Time Machine <http://timesmachine.nytimes.com/timesmachine/1979/07/21/111186136.html?zoom=15&pageNumber=6> (NYT login required) or from the authors upon request.

5. The data were originally collected by Frank R. Baumgartner and Bryan D. Jones, with the support of National Science Foundation grant number SBR 9320922 0111611, and were distributed through the Department of Government at the University of Texas at Austin. Neither NSF nor the original collectors of the data bear any responsibility for the analysis reported here.

6. We included the following hearing codes: 200 (General Civil Rights, Minority Issues and Civil Liberties), 201 (Ethnic Minority and Racial Group Discrimination), 206 (Voting Rights and Issues), 207 (Freedom of Speech and Religion), 209 (Anti-Government Activities), and 299 (Other). We removed all hearings and laws related to other ethnic minority groups (e.g., hearings focused on discrimination of Japanese Americans, the disabled, sex, and age discrimination).

7. The availability of U.S. newspaper indices is covered in Fritze, Coutts, and Vyhnanek (2004, pp. 102–105). In summary, while other major U.S. newspaper indices exist, the *New York Times* is the only index that exists prior to 1972, itself going back in index format to 1851. Thus, it is the only available newspaper of record in the United States that would enable index-based research on most-modern social movements, as collective action in post-WWII America began fomenting in the 1940s for the Civil Rights movement, and in the 1960s for most other movements.

8. To help track differences between print and digital newspaper sources – including changes in online newspaper editions as they evolve over time – the

website for NewsDiffs tracks online news over time, archiving changes in online editions/articles after publication: <http://newsdiffs.org>. NewsDiffs currently tracks the [nytimes.com](http://nytimes.com), [cnn.com](http://cnn.com), [politico.com](http://politico.com), [washingtonpost.com](http://washingtonpost.com), and [bbc.co.uk](http://bbc.co.uk)

9. We were able to offer one comparison case for generalizability, examining protest events organized by the U.S. Environmental Movement using data from Jenkins' same *NYT Index* coding project and the DCA full-story data. We find that the capture rate of the NYT Index protest events for the Environmental Movement to be 38.3% from 1960 to 1995 (88 index events to 230 DCA events). This compares to an index capture rate of 31.6% for the Civil Rights movement protest events over the same time period (1,239 index events to 3,919 DCA events), as discussed in detail on page 11.

## REFERENCES

- Agnone, J. (2007). Amplifying public opinion: The policy impact of the U.S. environmental movement. *Social Forces*, 85(4), 1593–1620.
- Althaus, S., Edy, J., & Phalen, P. (2001). Using substitutes for full-text news stories in content analysis: Which text is best? *American Journal of Political Science*, 45(3), 707–723.
- Amenta, E., Halfmann, D., & Young, M. P. (1999). The strategies and contexts of social protest: Political mediation and the impact of the Townsend Movement In California. *Mobilization: An International Journal*, 4(1), 1–23.
- Andrews, K. T. (2001). Social movements and policy implementation: The Mississippi civil rights movement and the war on poverty, 1965 to 1971. *American Sociological Review*, 66, 71–95.
- Baumgartner, F. R., & Jones, B. D. (1993). *Agendas and instability in American politics*. Chicago, IL: University of Chicago Press.
- Baumgartner, F. B., & Jones, B. D. (2002). *Policy dynamics*. Chicago, IL: University of Chicago Press.
- Beyerlein, K., Soule, S. A., & Martin, N. (2015). Prayers, protest, and police: How religion influences police presence at collective action events in the United States, 1960 to 1995. *American Sociological Review*, 80(6), 1250–1271.
- Burstein, P. (1985). *Discrimination, jobs, and politics*. Chicago, IL: University of Chicago Press.
- Davenport, C., Soule, S. A., & Armstrong II, D. A. (2011). Protesting while black? The differential policing of American activism, 1960–1990. *American Sociological Review*, 76(1), 152–178.
- Earl, J., Martin, A., McCarthy, J. D., & Soule, S. (2004). The use of newspaper data in the study of collective action. *Annual Review of Sociology*, 30, 65–80.
- Earl, J., & Soule, S. A. (2010). The impacts of repression: The effect of police presence and action on subsequent protest rates. *Research in Social Movements, Conflicts and Change*, 30, 75–113.
- Earl, J., Soule, S. A., & McCarthy, J. D. (2003). Protest under fire? Explaining protest policing. *American Sociological Review*, 69, 581–606.
- Eyerman, R., & Jamison, A. (1991). *Social movements: A cognitive approach*. University Park, PA: Pennsylvania State University Press

- Fritze, R. H., Coutts, B. E., & Vyhnanek, L. A. (2004). *Reference sources in history, an introductory guide* (2nd ed.). Santa Barbara, CA: ABC-CLIO.
- Herman, E. S., & Chomsky, N. (1988). *Manufacturing consent*. New York, NY: Pantheon.
- Hopkins, D. J., & King, G. (2010). A method of automated nonparametric content analysis for social science. *American Journal of Political Science*, *54*(1), 229–247.
- Jenkins, J. C., & Eckert, C. M. (1986). Channeling black insurgency. *American Sociological Review*, *51*, 812–829.
- Jenkins, J. C., Jacobs, D., & Agnone, J. (2003). The development of civil rights protest, 1947–1997: Collective grievances, mobilizing structures and political opportunities. *American Journal of Sociology*, *109*, 277–303. (Reprinted in Doug McAdam and David Snow (Eds.), *Readings on Social Movements: Origins, Dynamics, and Outcomes*, Oxford University Press 2009).
- Jenkins, J. C., & Maher, T. V. (2016). What should we do about source selection in event data: Challenges, progress and possible solutions. *International Journal of Sociology*, *46*, 1–16.
- Jenkins, J. C., & Perrow, C. (1977). Insurgency of the powerless: Farm worker movements (1946–1972). *American Sociological Review*, *42*(2), 249–268.
- Johnson, E. W. (2008). Social movement size, organizational diversity and the making of federal law. *Social Forces*, *86*(3), 967–993.
- Johnson, E. W., Agnone, J., & McCarthy, J. D. (2010). Movement organizations, synergistic tactics and environmental public policy. *Social Forces*, *88*(5), 2267–2292.
- Jung, W., King, B. G., & Soule, S. A. (2014). Issue bricolage: Explaining the configuration of the social movement sector, 1960–1995. *American Journal of Sociology*, *120*(1), 187–225.
- King, B. G., Bentele, K. G., & Soule, S. A. (2007). Protest and policymaking: Explaining fluctuation in congressional attention to rights issues, 1960–1986. *Social Forces*, *86*(1), 137–164.
- Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage.
- Martin, A. W., McCarthy, J. D., & McPhail, C. (2009). Why targets matter: Toward a more inclusive model of collective violence. *American Sociological Review*, *74*(5), 821–841.
- McAdam, D. (1982). *Political process and the development of black insurgency, 1930–1970*. Chicago, IL: University of Chicago Press.
- McAdam, D. (1983). Tactical innovation and the pace of insurgency. *American Sociological Review*, *48*, 735–754.
- McAdam, D., & Su, Y. (2002). The war at home: Antiwar protests and congressional voting, 1965–1973. *American Sociological Review*, *67*, 696–721.
- McCarthy, J. D., McPhail, C., & Smith, J. (1996). Images of protest: Dimensions of selection bias in media coverage of Washington demonstrations, 1982 and 1991. *American Sociological Review*, *61*, 478–499.
- Meyer, D. S., & Minkoff, D. C. (2004). Conceptualizing political opportunity. *Social Forces*, *82*(4), 1457–1492.
- Myers, D. J., & Caniglia, B. S. (2004). All the rioting that's fit to print: Selection effects in national newspaper coverage of civil disorders, 1968–1969. *American Sociological Review*, *69*(4), 519–543.
- New York Times. (1960–1995). New York Times Index. *New York Times*.



- Nicholson, B. (2013). The digital turn: Exploring the methodological possibilities of digital newspaper archives. *Media History*, 19, 59–73.
- Oliver, P. E., & Myers, D. J. (1999). How events enter the public sphere: Conflict, location, and sponsorship in local newspaper coverage of public events. *American Journal of Sociology*, 105, 38–87.
- Olzak, S. (1989). Analysis of events in the study of collective action. *Annual Review of Sociology*, 15, 119–141.
- Olzak, S., & Uhrig, S. C. N. (2001). The ecology of tactical overlap. *American Sociological Review*, 66(5), 694–717.
- Olzak, S., & Soule, S. A. (2009). Cross-cutting effects of protest, congressional hearings and bills on the environment. *Social Forces*, 88(1), 201–226.
- Ortiz, D. G., Myers, D. J., Walls, N. E., & Diaz, M. E. D. (2005). Where do we stand with newspaper data? *Mobilization*, 10(3), 397–420.
- Ridout, T. N., Fowler, E. F., & Searles, K. (2012). Exploring the validity of electronic newspaper databases. *International Journal of Social Research Methodology*, 15(6), 451–466.
- Ring-Ramirez, M., Reynolds-Stenson, H., & Earl, J. (2014). Culturally constrained contention: Mapping the meaning structure of the repertoire of contention. *Mobilization*, 19(4), 489–505.
- Soule, S. A., & King, B. G. (2008). Competition and resource partitioning in three social movement industries. *American Journal of Sociology*, 113(6), 1568–1610.
- Spilerman, S. (1970). The causes of racial disturbances: A comparison of alternative explanations. *American Sociological Review*, 35, 627–649.
- Stimson, J. A. ([1991] 1999). *Public opinion in America: Moods, cycles, and swings*. Boulder, CO: Westview Press.
- Tarrow, S. (1989). *Democracy and disorder: Protest and politics in Italy, 1965–1975*. New York, NY: Oxford University Press.
- Taylor, C. L., & Jodice, D. A. (1983). *The world handbook of social and political indicators* (Vol. 3). New Haven, CT: Yale University Press.
- US Bureau of the Census. (1980). *Conference on Census Undercount: Proceedings of the 1980 Conference*. Washington, DC: U.S. Department of Commerce.
- Van Dyke, N., Soule, S. A., & Taylor, V. A. (2004). The targets of social movements: Beyond a focus on the state. In D. J. Myers & D. M. Cress (Eds.), *Authority in contention* (Vol. 25, pp. 27–51). *Research in Social Movements, Conflict and Change*. Bingley, UK: Emerald Group Publishing Limited.
- Walker, E. T., Martin, A. W., & McCarthy, J. D. (2008). Confronting the state, the corporation, and the Academy: The influence of institutional targets on social movement repertoires. *American Journal of Sociology*, 114(1), 35–76.
- Wang, D. J., & Soule, S. A. (2012). Social movement organizational collaboration: Networks of learning and the diffusion of protest tactics, 1960–1995. *American Journal of Sociology*, 117(6), 1674–1722.
- Weiner, M. F. (2011). All the news that's fit to print? Silence and voice in mainstream and ethnic press accounts of African American protest. *Research in Social Movements, Conflict and Change*, 31, 297–324.
- Woolley, J. T. (2000). Using media-based data in studies of politics. *American Journal of Political Science*, 44(1), 156–173.

## APPENDIX

**Table A1.** Variable Descriptive Statistics, 1961–1995.

	Mean	Std. Deviation	Minimum	Maximum	<i>N</i>
<i>Dependent variables</i>					
Protests (index)	30.8	47.7	1	240	35
Protests (full text)	100.3	130.9	12	578	35
<i>Independent variables</i>					
Black/White income gap	0.6	0	0.5	0.6	35
NAACP membership (LN)	6	0.1	5.9	6.2	35
Northern democrats in congress	15.6	18.3	0	40	35
Divided congress	0.6	0.5	0	1	35
Black congressmen	16.9	9.8	3	40	35
Vietnam war deaths (LN)	2.5	3.5	0	9.4	35
Black unemployment	12	3.2	6.4	19.5	35
Civil rights laws ( $t-1$ )	0.9	1	0	4	35
Civil rights hearings ( $t-1$ )	12.8	5.8	3	21	35

**Table A2.** Pearson Correlation Coefficients, 1961–1995.

	Protests (Index)	Protests (Full Text)	Public Opinion ( <i>t</i> -1)	Republican Incumbent President	Black/ White Income Gap	NAACP Membership (LN)	Northern Democrats in Congress	Divided Congress	Black Congressmen	Black Vietnam War Deaths (LN)	Black Unemployment	Civil Rights Laws ( <i>t</i> -1)	Civil Rights Hearings ( <i>t</i> -1)
Protests (index)	1												
Protests (full text)	0.875	1											
Public opinion ( <i>t</i> -1)	0.52	0.631	1										
Republican incumbent president	-0.201	-0.178	0.002	1									
Black/White income gap	-0.468	-0.514	-0.225	0.027	1								
NAACP membership (LN)	0.441	0.645	0.4	-0.136	-0.035	1							
Northern democrats in congress	0.435	0.445	0.186	-0.352	-0.286	0.352	1						
Divided congress	-0.481	-0.501	-0.285	0.333	0.376	-0.383	-0.956	1					
Black congressmen	-0.606	-0.652	-0.506	0.124	0.298	-0.482	-0.204	0.349	1				
Black Vietnam war deaths (LN)	0.599	0.63	0.687	-0.17	0.057	0.507	0.261	-0.322	-0.68	1			
Black unemployment	-0.444	-0.481	-0.706	0.168	-0.196	-0.438	-0.318	0.309	0.364	-0.792	1		
Civil rights laws ( <i>t</i> -1)	-0.193	-0.26	-0.121	0.024	-0.057	-0.277	-0.102	0.134	0.289	-0.245	0.073	1	
Civil rights hearings ( <i>t</i> -1)	-0.582	-0.564	-0.56	0.173	0.123	-0.46	-0.404	0.423	0.607	-0.605	0.49	0.235	1

**Table A3.** Effects of a Data Collection Method on Poisson Models Predicting Civil Rights Protest Activity, 1961–1995.

Variables	Model 1 Index	Model 2 Full-story	Model 3 Index	Model 4 Full-story	Model 5 Index	Model 6 Full-story	Model 7 Index	Model 8 Full-story
Republican incumbent president	-0.425** (0.187)	0.0318 (0.122)	-0.356** (0.180)	0.0256 (0.116)	-0.395** (0.193)	-0.0391 (0.104)	-0.330 (0.231)	-0.00254 (0.116)
Black/White income gap	-17.11*** (5.739)	-15.58*** (1.872)	-14.56*** (3.672)	-16.91*** (2.000)	-13.75*** (4.231)	-15.61*** (1.526)	-14.63*** (4.829)	-15.82*** (1.676)
NAACP membership (LN)	-0.519 (1.768)	2.761** (1.205)	-0.987 (1.533)	3.234*** (1.077)	-0.989 (1.513)	3.198*** (1.031)	-1.431 (1.551)	2.930*** (1.004)
Northern democrats in congress	0.0504* (0.0295)	0.0227* (0.0122)	0.0649** (0.0266)	0.0246* (0.0134)	0.0749** (0.0320)	0.0400*** (0.00953)	0.0787*** (0.0304)	0.0401*** (0.00986)
Divided congress	2.059* (1.165)	1.148** (0.459)	2.393** (1.086)	1.149** (0.503)	2.720** (1.216)	1.690*** (0.364)	2.719** (1.167)	1.636*** (0.350)
Black congressmen	-0.0626** (0.0283)	-0.0557*** (0.0160)	-0.0986*** (0.0318)	-0.0560*** (0.0139)	-0.0947*** (0.0269)	-0.0484*** (0.0112)	-0.0947*** (0.0249)	-0.0495*** (0.0104)
Vietnam war deaths (LN)	0.137* (0.0780)	0.0565 (0.0555)	0.136** (0.0560)	0.106** (0.0429)	0.153*** (0.0479)	0.129*** (0.0426)	0.190*** (0.0526)	0.140*** (0.0400)
Black unemployment	-0.0644 (0.0606)	-0.108** (0.0480)	-0.0424 (0.0556)	-0.0763* (0.0416)	-0.00738 (0.0755)	-0.0239 (0.0467)	0.00480 (0.0690)	-0.0253 (0.0435)
Protests ( $t-1$ )			-0.00519*** (0.00134)	-0.00109** (0.000481)	-0.00565*** (0.00173)	-0.00114*** (0.000415)	-0.00590*** (0.00154)	-0.000959** (0.000386)
Public opinion ( $t-1$ )					0.0220 (0.0304)	0.0356** (0.0148)	0.0239 (0.0295)	0.0340** (0.0138)
Civil rights hearings ( $t-1$ )							0.0284 (0.0246)	0.0185* (0.0107)

Constant	15.15 (11.54)	-2.644 (7.150)	16.57 (10.09)	-5.135 (6.716)	13.96 (11.17)	-9.126 (6.354)	16.34 (12.09)	-7.528 (6.210)
Observations	35	35	35	35	35	35	35	35
Pseudo $R^2$	0.798	0.896	0.834	0.906	0.836	0.911	0.841	0.914
Log likelihood	-179.2	-226.1	-146.8	-204.0	-145.6	-193.1	-141.2	-185.9

---

Robust standard errors in parentheses.

\*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ .