WHERE DO WE STAND WITH NEWSPAPER DATA? [*]

David G. Ortiz, Daniel J. Myers, N. Eugene Walls, and Maria-Elena D. Diaz[†]

We review the literature that identifies selection bias in media-based data and propose a theo-retical model of the sources of these biases. Given the implications for validity and reliability, we conclude that newspaper data often do not reach acceptable standards for event analysis and that using them can distort findings and misguide theorizing. Furthermore, media selection biases are resistant to correction procedures largely because they are unstable across media sources, time, and location. We end with a plea for more circumspect approaches to media data that fully and openly consider the implications of their inherent limitations.

Measuring protest is a complicated venture, one that researchers have wrestled with for many years. Even a seemingly straightforward question such as, "How much protest happened last year?" can be difficult to answer in any definitive way because researchers cannot directly collect this information for any social unit of substantial size. We depend, therefore, on the work of others to help us track protest events.

More often than not, we turn to newspaper records to find protest events, to identify what occurred during the event, and to determine what the event was about. Because newspaper data are relatively easy to collect and are sometimes the only continuously available source of event data, they have been used by social scientists to analyze a wide range of activity including protest and collective violence (Gurr 1968; Jenkins and Eckert 1986; Jenkins and Perrow 1977; Kriesi, Koopmans, Duyvendak, and Guigni 1995; Lieberson and Silverman 1965; McAdam 1982; Myers 1997, 2000; Olzak 1992; Shorter and Tilly 1974; Spilerman 1970, 1976), assas-sinations, intra- and interstate wars, coups and revolutions, state repression (Taylor and Jodice 1983; Paige 1975; Rummel 1966), elections (Banks 1997), management control of corporations (Burch 1972), long-term changes in public opinion (Galambos 1975), strikes (Shorter and Tilly 1974), public signals about monetary policy (Woolley 2000), and media discourse and practices (Koopmans and Olzak 2004; Koopmans and Statham 1999; McAdam and Su 2002; Ferree 2003; Barnett and Woywode 2004).

Unfortunately, newspaper content is not created for the purpose of conducting social scientific research nor is it intended to capture or sample all protests or other political events, even in a limited geographic area. As a result, we must ask questions about the reliability and validity of newspaper data.

When newspaper event data came into wide use, a series of studies began to ask critical questions about the role of media processes in producing the data (e.g., Danzger 1975; Snyder and Kelly 1977; Jackman and Boyd 1979; Taylor and Hudson 1972; Hazelwood and West 1974;

Franzosi 1987). The concerns represented in these early pieces came under new scrutiny when researchers began to compare newspaper data to police and other governmental records (McCarthy, McPhail, and Smith 1996), which in turn unleashed a whole wave of studies, conferences, and volumes dedicated to understanding the selection biases inherent in newspaper and other media data (Barranco and Wisler 1999; Koopmans 1995; Hocke 1998; Mueller 1997a, 1997b; Myers and Caniglia 2004; Oliver and Myers 1999; Oliver and Maney 2000; Rucht, Koopmans, and Neidhardt 1995). These and other studies clearly documented that newspapers are not a transparent conduit of information about protest, and that important systematic selection factors and processes affect the types of data available in newspapers for social scientists.

The reaction of scholars to these methodological studies has been mixed. Some have completely ignored selection issues and proceeded as if newspaper data were a random sample of the universe of events. Others have concluded that newspaper data are not critically flawed and that, with minimal care, such data can be used without producing serious distortion in our analyses. Yet others have decided that the shortcomings are relatively serious and have proposed ways of correcting for the distortions produced by selection processes. Given these disparate approaches, an assessment of where we stand with newspaper data is in order: Based on the record of media-selection studies, can newspaper data be considered reasonably valid and reliable?

Although a recent review of this question by Earl, Martin, McCarthy, and Soule (2004) concluded that newspaper data were largely sound and that researchers could proceed with caution, our read of the record is much less optimistic. We find that the body of literature has located very serious flaws in newspaper data and suggests even more problems than have yet been fully documented. Furthermore, we find that the correctives that have been offered are inadequate and could potentially make things worse. We conclude that researchers who wish to continue using event data based on newspaper and other mass media records need to take the shortcomings of the data much more seriously, and need to spend more time and print space thinking about how these shortcomings may affect their results, conclusions, and theorizing.

THE EMPIRICAL RECORD ON NEWSPAPER DATA

Since the early 1970s, sociologists and political scientists have been publishing findings about the usefulness and accuracy of newspaper data. The central area of concern has been what most often has been called "selection bias." In the present context, selection bias refers to how events are either selected into the media record or left unreported. In contrast, "description bias" refers to how events are represented in the media record, given that they are reported. In this article, we focus on the issue of selection bias.

A large number of factors that affect selection bias have been identified in prior studies and we organize them into four clusters. Admittedly, these clusters are somewhat artificial and other methods of grouping the findings could have been adopted, but this particular scheme produces a framework to organize the findings, building toward a model of media selection that we discuss later in the paper. First, we examine event characteristics that affect coverage. Second, we identify influential contextual factors. Third, we discuss media structure and characteristics of media outlets. Finally, the fourth cluster contains factors related to research processes—biases created by researchers as they collect and process the media record.

Event Characteristics

Event Intensity. The most consistent finding in the media bias literature is that the more intense an event, the more likely it is to be covered by newspapers. Event intensity has been operationalized in a number of ways, the most common being crowd size. All other things held constant, larger events are significantly more likely to be reported than are smaller events (McCarthy et al. 1996; Barranco and Wisler 1999; Oliver and Myers 1999; Oliver and Maney

2000). But many other measures of intensity have been identified as well. Most prominent among these are event duration, presence of violence including injuries and deaths, presence of police or other repressive social agents, number of arrests, and property damage (e.g., Snyder and Kelly 1977; Baranco and Wisler 1999; Myers and Caniglia 2004). Whether used as single indicators or combined into an index, these intensity indicators strongly increase the chances that the event receives newspaper coverage.

Conflict. Some intensity indicators are also measurements of conflict between one or more parties. However, the presence of conflict in an event has an independent and direct effect on news coverage. Generally speaking, conflictual events receive more coverage than non-conflictual events (Oliver and Myers 1999). Thus, intensity indicators such as violence, the presence of police, and targeted property damage may further increase coverage because they indicate that disputants have come into conflict. Likewise, the presence of counterdemonstrators at an event embodies conflict and sets the stage for the possibility of emotional and physical clashes between opposing sides of an issue (Oliver and Maney 2000).

Sponsorship. Events sponsored by organizations have increased chances of media coverage (Oliver and Maney 2000). Organizational sponsorship brings not only credibility and legitimacy but also resources, infrastructure, networks, and experience. If an event's organizers are familiar with the routines and priorities of a news outlet, they can target particular news gatherers to increase the likelihood of coverage (Gamson and Wolfsfeld 1993). In fact, research has found that organizers who do not play this "insider media game" (Ryan 1991) are at a significant disadvantage.

Significance of Actors Involved. If a newspaper's staff or audience perceives that actors involved in an event are politically or culturally significant, then those events are more likely to be reported (Snyder and Kelly 1977). This significance may arise due to a social actor's status, notoriety, or celebrity. For example, Myers and Caniglia (2004) found that events involving secondary school students were less likely to be reported, while events involving college students received greater attention. Significance of actors does not necessarily mean positive attitudes toward the actors, and there is evidence that negative notoriety of social actors (Richards and McCarthy 2002) can also drive up coverage. By the same token, the media may completely fail to cover events regarding certain populations or topics because they deem them unimportant, distasteful, or not of interest to their readership (Harrison 2001).

Locale. An event's location can matter in several ways. In the following section we will discuss location in spatial relation to media sources and to other contextual elements, but location can also be considered an event characteristic. Activists can choose or target specific locales for their events because they hold a special meaning for them, the counterprotestors, or the state. As such, these locales can add to the drama and theatre of the protest events, thereby garnering additional media attention (Oliver and Myers 1999).

Contextual Factors

Characteristics of protest events go a long way toward determining whether or not they will receive coverage, but events do not occur in a vacuum and certain elements of the context can also have a powerful effect on the chances of coverage. These contextual elements include event density, proximity to media sources, and characteristics of the place where the event occurs.

Event Density. One key contextual element that influences coverage is the recent occurrence of other events. Protest events must compete with other protests and other kinds of news for coverage. This is sometimes called the "news hole" effect. Although news holes are elastic to some degree, a smaller proportion of events will likely be reported when there is a great deal of newsworthy activity, compared to times when fewer events take place (Hocke 1998; Oliver and Myers 1999; Oliver and Maney 2000). This event density effect has been operationalized in various ways, such as counts of recent events (Myers and Caniglia 2004), or by identifying slow

news days, like Mondays, and comparing them with heavier days, like Fridays (Oliver and Myers 1999).

Proximity. In addition to the symbolic function of an event's location, place also matters based on its distance from the media source that may record the event (Snyder and Kelly 1977; McCarthy et al. 1996; Mueller 1997a). Media sources are less likely to attend to and report events that are far away due to a general lack of awareness of the event, lack of staff to cover the event, or a perception that the event is not of interest to the newspaper's audience. These effects can be extremely powerful. Myers and Caniglia (2004), for example, found that the *New York Times* was thirty times more likely to report events occurring in New York City compared to those occurring elsewhere. Advances in communication technology may be reducing both the infrastructure difficulties and the localized audience demand that drive proximity effects, but how much change is occurring is an open question.

Location Characteristics. Media gatekeepers often consider some cities and local environments more important than others. Protests that occur in large, urban areas may be considered more newsworthy because these cities carry more political and cultural clout (Myers 2000). Moreover, as Myers and Caniglia (2004) demonstrated, the media may, at times, view urban areas with specific characteristics as more important. Due to the salience of race relations, the U.S. media in the 1960s seems to have focused more on cities with larger African American populations. Whether the media viewed urban rioters as political actors with legitimate grievances or as criminals and hooligans (Jefferies, Turner, and Morris 1971), they were nonetheless important newsmakers. More recently, we might find that the media focused inordinate attention on urban areas perceived to have large lesbian and gay populations (such as San Francisco, New York City, and Los Angeles) during the same sex marriage protest events of 2004-2005.

Media Structure and Media Outlets

Beyond events and their location in space and time, selection is also a function of the overall structure of the mass media and the specific organizational characteristics of individual media outlets. As an embedded component of the larger social structure, the media system is heavily influenced by its corporate and financial interests, by its intended audience, and by the political climate in which it exists. Here we identify three specific factors: profit motive/corporate interests, attention cycles, and political climate. In addition, individual media outlets have specific mechanics of collecting information and self-definitions (rooted in their intended audience) that further contribute to selection.

Profit Motive and Corporate Interests. News media is big business in a capitalist economy. Consequently, profit motives play a major role in determining news coverage. A variety of studies across disciplines have established that news that threatens the flow of profits or challenges the existing power structure is less likely to be covered (Bagdikian 2000; Parenti 1993; Gamson, Croteau, Hoynes, and Sasson 1992; Lee and Solomon 1990; Herman and Chomsky 1988; Gitlin 1980; MacGregor 1997; Smith, McCarthy, McPhail, and Augustyn 2001). Beyond direct influence on event selection, corporate interest can also affect the framing of those events that are reported, promote an attention cycle for a "hot" issue, or exclude voices of newsmakers who challenge corporate agendas and authority. Advertisers, armed with financial incentives, can also persuade media outlets to avoid or emphasize particular perspectives. Moreover, the concentration of media ownership that started in the 1970s and continues today exacerbates the problem by limiting the diversity of news and perspectives communicated by the media (Herman 1995).

Attention Cycles. Media attention cycles refer to the phenomenon whereby certain issues gain prominence in the media for a period of time before fading into obscurity (McCarthy, et al. 1996; McCombs and Shaw 1972). In the midst of an issue cycle, reporters and editors select stories connected to the issue, which can lead to over-representation of related events. McCarthy and colleagues (1996) analyze the protest coverage of newspapers and news stations in

Washington, D.C. for two different time periods and find evidence to suggest that media attention cycles are the second strongest predictor of protest reporting, after demonstration size.

While the attention cycle idea is plausible and supported by some suggestive evidence, attention cycles can be difficult to identify outside of the very newspaper record that they are supposed to affect. In addition, it is important to note that evidence suggests that attention cycles are inconsistent across issues, time, space, and newspaper sources (McCarthy et al. 1996). Roberts, Wanta, and Dzwo (2002) also found that lags between media coverage and public concern differed by issue. At present, then, the evidence about the presence and nature of attention cycles is equivocal.

Political Climate. When authoritarian regimes are in power, it is obvious that censorship will alter news-reporting practices (Snyder and Kelly 1977). However, in less repressive states, where the political climate can vary across time and location, more subtle influences on news reporting are usually at work. Political elites have inordinate access to the mass media, and can champion particular issues and events or dissuade coverage of less-favored actors and events (Lipksy and Olson 1976; Goren 1980; Shoemaker 1988). Barranco and Wisler (1999), for example, found that a conservative newspaper under-reported violent demonstrations and suggested that the paper succumbed to political pressure intended to limit diffusion of these events. Other researchers have found similar politically influenced attempts to limit copycat behavior by squelching mass media coverage (Lipsky and Olson 1976; Monroy and Myers 2003; National Advisory Commission 1968). Furthermore, competing activity in the political field has also been shown to reduce coverage of protest at certain times (Oliver and Maney 2000; Meyer 1993, 1995; Olzak 1992) and increase it at others (Tilly 1997).

Media Mechanics. Media outlets are limited partly in what they report by the routines, infrastructure, and resources they use to collect information. Danzger (1975) was the first to call attention to this media bias issue by showing that the numbers of civil disorders reported for a city was strongly correlated with the presence of a wire service office in that city. Since then, scholars have documented a number of other important mechanical features of the reporting process that have serious effects on event coverage. For example, reporters have beats and routines that place certain events in their paths, while others escape their attention because they are too far removed from the usual ways of conducting reporters' jobs (Oliver and Myers 1999). This same dynamic increases the chance of coverage for routinized events that conform to expectations about activities likely to occur at anticipated dates, times, and places (Oliver and Myers 1999; Oliver and Maney 2000).
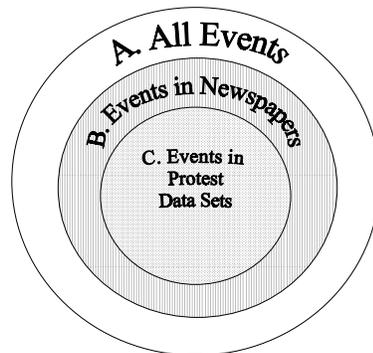
Reporters often turn to familiar and easily accessible government officials for commentary on developing events (Gans 1979). Events that allow such contributions increase their chances of coverage and at the same time limit information from other sources (Gamson et al. 1992; Gitlin 1980; Oliver and Maney 2000). Newspapers also depend heavily on centralized electronic resources for stories. As this trend has grown, local news staffs have been downsized which in turn reduces the coverage of certain types of events—among them social movement activities (Herman 1995).

Audience Characteristics and Self-definition. Newspapers and periodicals can be characterized as certain *types* of publications based on what kinds of events they focus on and what audience they intend to reach. The geographic scope of distribution (local, regional, national) defines an audience and, therefore, what events may be of interest to them (Myers and Caniglia 2004). Leftist newspapers are more likely to cover social movement events (Franzosi 1987; Kriesi et al. 1995), and liberal newspapers are more likely to cover protest events than conservative newspapers (Oliver and Myers 1999). Periodicals may also target certain demographic groups (such as young adults, a particular racial group, or women), which will affect the types of stories that they are likely to cover. Finally, circulation numbers also have a direct effect on the physical size of the newspaper, thereby decreasing the space for coverage if circulation is small, or increasing it as circulation grows.

Research Processes

Bias in newspaper-derived data does not end with media practices. More selection of events and information about them occurs after the media record is produced. Only in rare cases can social scientists collect information from all newspapers that might have covered the events of interest (Jackman and Boyd 1979; Koopmans 1995). Instead, scholars usually extract information by using some kind of sampling technique that will reduce the time and cost of data collection. Thus, there are two steps in the selection process that may produce bias. First, within the universe of relevant events (area A in figure 1), there is a subset reported by at least one newspaper (area B). These are the selection processes we have discussed up to this point. Second, among those events appearing in the media record (area B), there is an even smaller subset of events (area C) that actually appears in social science data. Although this second selection process has been examined less, it is just as important as the first because when constructing their data, scholars do not attempt to draw a random sample of events from the overall media record. In other words, the events in area C are unlikely to be fully representative of the events in area B, and even less so of events in area A.

Figure 1. Two Levels of Selection Bias

Limiting Sources. One way in which scholars have addressed the time and cost issues associated with collecting newspaper data is to limit the number of media sources from which data will be extracted. Most often this means collecting data from only one or two sources and, in many cases, just the *New York Times*. Studies on the validity and reliability of the *Times* demonstrate that it is the best single newspaper source for political event data, reporting more events and providing more detail than any other singular newspaper source (Taylor and Jodice 1983; Jackman and Boyd 1979; Paige 1975; Olzak and Olivier 1994; Olzak 1982, 1992). But even using the best newspaper does not necessarily produce a representative sample.

Some attempts to study this problem have concluded that adding one or two additional sources does not substantively change conclusions drawn from the *Times* data (Jackman and Boyd 1979; Jenkins and Perrow 1977), and because the costs involved in adding even one additional source are substantial, these findings have led to a recommendation against supplementing the *Times* with additional sources. Other researchers disagree with this approach. Woolley (2000: 161) shows that coverage of child abuse in five different media outlets differs substantially: "Our conclusions about the shape of the trend and about variability around the trend depend on the newspaper." Also, Myers and Caniglia (2004) compare the coverage of collective violence in the *New York Times* and the *Washington Post* to the coverage carried by all daily newspapers in the United States. Their findings reveal serious, significant selection effects in the *Times* and *Post* that substantially distort analyses and the subsequent conclusions. This leads us to believe that the selection problems with single, national media sources—even the *best* national source—may be more problematic than previously thought.

A potential solution to the single-source problem is to add additional newspapers. Even though this strategy may increase reliability, it may come at a high cost and does not necessarily guarantee that the data will improve as a result. Myers and Caniglia (2004), for example, show that in certain circumstances, adding the *Washington Post* to the *Times* actually increased discrepancies in the data.

Systematic Day Sampling Techniques. Another technique is to sample based on days of the week. If non-violent protest tends to be concentrated on the weekends, for example, Monday newspapers will report the bulk of events (Kriesi et al. 1995; Barranco and Wisler 1999). Unfortunately, not all protest is concentrated on the weekends. Thus, data collected using this technique will relatively underrepresent events like labor strikes and student demonstrations that tend to occur during the week (Rucht and Ohlemacher 1992; Koopmans and Rucht 2002), and collective violence that is more evenly distributed across the week (Myers and Caniglia 2004).

Electronic Media Records. One way to circumvent the single-source problem may be the use of more recent electronic, searchable media records. In theory, it should be easier and cheaper to add many media sources to the data collection enterprise through electronic archives and clearinghouses. Yet at this point in time, scholars have found it difficult to work with these sources (Wagers 1992; Woolley 2000). For one, the content covered by electronic sources is constantly changing as lawsuits, copyright concerns, and storage capacity expand and contract the available information. Likewise, it can be very difficult to find search strings that produce consistent output, do not miss many events, and yet reasonably limit "false positives" (Schrodt 1994; Maney and Oliver 2001).

In our own work with electronic media databases, we identified additional problems that have not yet been documented in the literature. Even the most "comprehensive" media databases do not cover all major newspapers. At the time of our searches (winter of 2005), Lexis-Nexis, for example, did not include material from the *Chicago Tribune*, and Newspaper Source did not contain material from the *Los Angeles Times*. Even in those e-databases that carried the same newspapers, the range of coverage might be dissimilar either in time period, or in selected content. These limitations force researchers who wish to consult a wide range of newspapers to search more than one electronic database, an enterprise also fraught with difficulty since search mechanisms vary and thus do not allow for consistent searching across databases. For example, some databases would allow the hyphenated "sit-in" as a search term and others would not. Some would allow complex combinations of search terms and limits, while others had much more restricted capacities.

Timing of Searches. The time within which investigators conduct a search can also affect outcomes. Some protest events are announced in the media in advance, and then are not covered after they occur. Other events – such as non-planned, spontaneous events—may only be covered after the event has taken place. Smaller events may be mentioned in thematic articles, which attempt to demonstrate connections across a series of events, although they may not have been reported at the time the event occurred. Still others may be covered on their anniversaries or on anniversaries of other pertinent events (e.g., *Roe v. Wade* decision, Stonewall Riot, September 11). Thus, coverage of certain types of events might be more easily missed because of temporal search limits (Woolley 2000; Maney and Oliver 2001).

Coding. Finally, after the newspaper articles have been identified, researchers still have to deal with the procedures for extracting information and turning it into analyzable data (Franzosi 2004). Coding errors are not unique to media-based data, of course, but can be aggravated in the media context. This is because reporters and editors have already coded media data. Researchers are not coding events; they are recoding previously coded data.

Each decision the researcher makes about coding can fundamentally change the data in systematic ways that may alter conclusions. What characteristics of events lead to inclusion in the data set? What about missing information? In Earl, Soule, and McCarthy (2003), for example, the researchers assume that if the article does not mention arrests or police barricades,

then no arrests were made and no barricades were used. These kinds of assumptions are common, yet we do not know how they influence the findings and interpretations thereof.

Intercoder reliability can also be difficult to achieve given that it varies significantly as a function of what is being coded. Searches for restricted variables, such as number of deaths or assassinations, typically produce very high intercoder reliability, but more complex and ambiguous variables, such as crowd size and repression, can yield very unreliable results. Taylor and Jodice (1983), for example, found that intercoder reliability on the frequency of coups was only 54 percent. While coding can be a cause of inaccuracies, it is also generally well-documented and, therefore, more easily examined for bias (Woolley 2000). In contrast to the media record itself, researchers can control the coding process, study it, and produce improvement, particularly if they have the resources to triangulate data with multiple sources and coders.
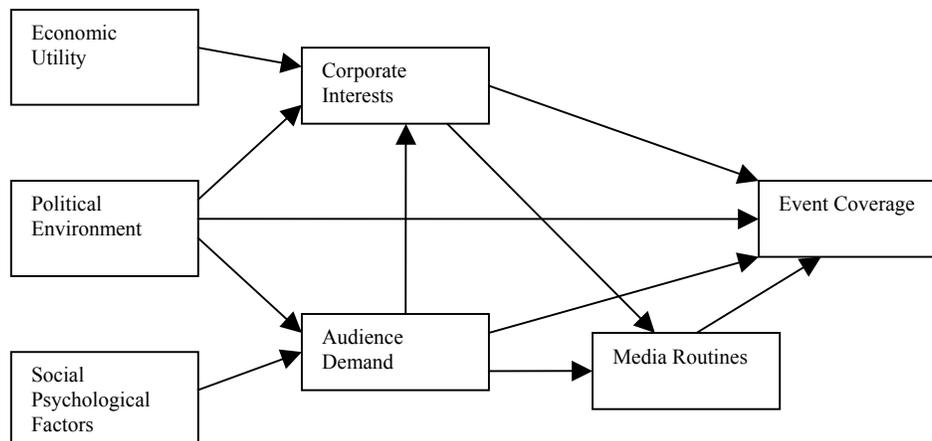
TOWARD A MODEL OF MEDIA SELECTION

As is clear from the foregoing review, researchers have identified a large number of factors that influence media coverage and can produce bias in empirical studies. Setting aside the research processes that contribute bias after the media record has been constructed, we have configured these findings into a model of media selection that understands the selection mechanisms as products of fundamental social processes. We are not, of course, the first to construct a model of media selection (e.g., Snyder and Kelly 1977 for an early example; and, Mueller 1997b for a more thorough treatment), but the model we discuss incorporates recent findings and attempts to trace the selection processes to fundamental theoretical constructs.

News outlets are profit-driven enterprises that must appeal to the audience in order to subsist in a particular political environment. Accordingly, it is necessary to firmly root a model of selection bias via theoretic notions of economic interests, political settings, and social psychological processes that give rise to the media production itself. These basic drivers produce corporate interests, audience demand, and media routines, which, in turn, produce the selection logic used to determine coverage (see figure 2).

This model reorganizes and reframes the main findings of bias studies in terms of the forces that motivate media production. In this way, it departs from previous models that only highlighted the immediate sources of selection bias and their effects on the data. By organizing the empirical findings of media-selection studies with these conceptual categories, the model

Figure 2. Media Selection Model

can flexibly inform future work on media production and can be used to guide researchers as they think through how media selection may bias their data and influence their findings.

Audience Demand

Audience demand plays a central role in the model. Many selection effects identified in past work can be tied directly to audience interest. For instance, the decline in coverage over geographic distance has often been treated as an infrastructure issue: It is more difficult to collect distant than nearby information. Yet even when collection difficulty is equalized (for example, by the Internet), geographic focus still remains. This occurs because newspapers have geographically concentrated audiences whose interests are locally focused.

But the question remains as to why audiences tend to be more interested in nearby events. The same question exists for the other audience-driven factors—like the novelty and boredom that is assumed to drive attention cycles, the draw to the drama of conflict and violence, and the cultural and symbolic significance of actors and locations. Why do these kinds of events draw attention while others do not? The answer exists in a parallel social psychological literature on the salience of a stimulus. Kiousis's (2004) review suggests two dimensions of salience, *attention* and *valence*, which would directly affect the coverage of specific events by defining the importance of the event to the audience.

1. Attention. This relates to those aspects of a stimulus that make it noticeable. Novelty, which makes the stimulus apparent in a sea of homogeneity, increases the likelihood that the stimulus will capture one's attention (Cooper, Burgoon, and Roter 2001). This is especially true if the novel behavior is also extreme and/or shocking (Cooper et al. 2001; Fiske 1980). Thus, unusually large events, events of longer duration, events with property damage, deaths, injuries or arrests, events of greater intensity, events where counterdemonstrators and/or significant actors are present, and unique event locations all increase the novelty of the event and, thus, its newsworthiness. The involvement of significant, familiar actors in an event also can trigger the "celebrity" effect, which has been shown to increase psychological arousal (Chen 2003) and influence the opinions and behaviors of individuals who identify with the actor (Basil 1996). This effect may be induced by either positive or negative notoriety.

Researchers have also found that individuals attend to stimuli that are personally relevant to them, inducing "utilitarian" motivation to watch news and cognitively process it (Perse 1992; Cooper et al. 2001). Because local environments are more personally relevant, salience of local news is higher, as demonstrated experimentally by Knobloch, Zillman, Gibson, and Karrh (2002). They found that regional ties in a story increased both the qualitative and quantitative knowledge of an event, while lack of regional ties resulted in either poorly processed knowledge about the event or the event being completely ignored. Relevance also produces cultural effects in determining audience salience because shared symbols and meaning-making processes make an issue or event more relevant. These shared cultural understandings can produce boundaries that define or at least influence relevance.

Attention effects are limited and fleeting, however. Individuals cannot attend to every personally relevant event, and novelty effects dissipate rapidly. Research shows, for example, that the public only attends to between five and seven issues at any one time (McCombs and Ghanem 2001; Shaw and McCombs 1977), which corresponds to psychological research that indicates that individuals can remember between five and nine topics at once (Klatzsky 1975). As new events arise and provide competition for old ones, the audience's attention to the old issues wanes rapidly.

2. Valence. Negatively valenced information carries more weight than positively valenced information (Leyens and Yzerbyt 1992), in part because it is rare and unexpected (Fiske 1980), and in part because it preempts other perceptual processes and is detected automatically (Pratto and Oliver 1991). Knobloch, Hastell, Zillman, and Callison (2003) conducted experiments in

which they manipulated the presence or absence of text-related images, and how threatening such images were to the reader. They found that the presence of any type of image had moderate effects in increasing attention to a specific article, but that the incorporation of threatening text-related images fostered both increased attention to the article and markedly more time spent reading it. Images and the affective tone of coverage shape both the audience perception of the events (Ghanem 1996; McCombs, Llamas, Lopez-Escobar, and Rey 1997) and the salience of the issue or event (Kiousis, Bantimaroudis, and Ban 1999). Thus media outlets can tap into audience interest and increase their sales by emphasizing negatively valenced events and aspects of their stories. Violence, deaths, injuries, property damage and/or the presence of counter-demonstrators fall into this emotionally laden category.

Corporate Interests

Audience demand is important because media outlets will respond to it as they attempt to maximize their own interests. In their pursuit of profit, news outlets attempt to maximize paid readership and advertising revenues. Thus, media gatekeepers attempt to select stories that will draw in new readers and retain their current reader base. Dramatic, novel, negatively valenced events are most likely to do this and these types of events appear in newspapers at higher rates than their less dramatic counterparts. But corporate interests are defined by more than just audience demands. News media operate in a political environment, one that can greatly influence their fortunes. Too much negative coverage of a political entity or office holder can reduce the newspapers access to resources and information or, in a more authoritarian context, can shut down the enterprise completely. Corporate interests and the interest structures of social and political elites directly affect decisions about newspaper content.[1]

Media Routines

Corporate interests and audience demand work together to define the everyday practice of identifying and reporting news. Because of corporate interests, reporters focus on identifying certain types of news and establish beats designed to capture those kinds of news. Likewise, media outlets establish routines and editorial decision-making standards based on what they perceive will be most important and interesting to their audience.

At the same time, some media routines exist simply because they are easier or more efficient to execute. Events that happen at the right time and place are more likely to be covered simply because they are more available to the reporting mechanisms. For example, activists routinely schedule protests to coincide with local news broadcasts, and by doing so make it easier for the news apparatus to access and broadcast their event. In this way, media routines can have an additional impact on coverage that is not fully dependent on corporate interests or audience interest.

Direct Effects: Politics, Interests, and Audience

While many of the effects in our model are indirect (audience demand works by defining corporate interests; political influence defines corporate interests and affects audience interest; corporate interests and audience demand shape media routines), each has its own direct effects on coverage as well. Although less common in open democratic societies where guarantees of press freedom are in place, the political structure can directly censor stories and keep them out of the press—even if audience demand and corporate interests would dictate otherwise. Audience demand can produce coverage of a story even if it requires stepping outside normal media production routines. Corporate interests can and have stepped in to squelch news stories even after they have been produced by the normal media routines.[2]

Although understanding the sources of media selection is important in itself, our more im-

mediate concern is methodological. Thus, having established a general model of media selection bias, we now turn our attention to the outcomes of this selection process on our research.

HOW BIG ARE THE PROBLEMS WITH NEWSPAPER DATA?

As scholars have studied newspaper data over the past decade, virtually everyone has come to agree that problems exist with the data. As evidenced by the long list of selection factors summarized above, media data are most certainly incomplete and biased to some degree. But how bad is it really? In the end, are the biases significant? Do they distort findings and misguide theorizing, or are they relatively minor and can be, with precautions, reasonably ignored? And even if the problems are substantial, can they perhaps be corrected to produce valid analyses?

The "how bad" question causes substantial disagreement among scholars who study and use newspaper data, yet it is particularly important for newspaper data (in contrast to other types of social science data) because scholars do not contribute to their design and mechanics of production. As Mueller (1997a: 821) put it, "In few other fields of the social sciences does the researcher have so little control over the data on which analysis depends." Analysts cannot intervene in newspaper reporting and editing practices to improve representativeness, change what characteristics of subjects (events) are recorded, or improve the consistency of data from event to event. We are, in a sense, stuck with what is handed to us in the media record.

Impressions of Protest

What kind of effects do these biases have on our understanding of protest and collective violence? In order to answer this question, we must first focus not on scholars, but on the general public's understandings of protest. When we rely on media data, and even more specifically on certain newspapers, we are reinforcing the media's understanding and presentation of culture and politics, and reproducing whatever biases they might have. When our studies treat *Times* data as scientific data, we implicitly place a stamp of authority on this information that is quite inconsistent with the unscientific way in which they were compiled. Given what we now know about these data, this is a fairly serious cause for concern.

Consider, for example, the Kerner Commission report's (National Advisory Commission on Civil Disorders 1968: 363) conclusion that "the portrayal of the violence . . . failed to accurately reflect its scale and character. The overall effect was, we believe, an exaggeration of both mood and event." When newspapers over-represent large, severe, and violent events (Myers and Caniglia 2004; McCarthy et al. 1996; Oliver and Myers 1999; Snyder and Kelly 1977; Barranco and Wisler 1999), they make dissidents appear more radical, more militant, and more outside the mainstream. This perception can lead to protesters being more easily dismissed, and greater support for heavy-handed repression. On the other hand, protest events that might be considered to be helpful contributions to the democratic process are relatively ignored. Conflicts in protest are also over-emphasized (Oliver and Myers 1999), suggesting that the inaccurate portrayal of protest can contribute to a sense of heightened racial animosity, for example. Whether social scientists have a normative obligation to avoid such incendiary portrayals is one question, but even if one takes a purely scientific stance toward the data, it is clear that these kinds of data practices make even our simple descriptive accounts of protest waves biased.

Newspaper and single source biases also tend to make events seem more clustered, both temporally (via attention cycles, McCarthy et al. 1996) and geographically (Myers and Caniglia 2004). This clustering has the effect of making the waves of activity seem more pronounced. It also exaggerates the seriousness of protest and collective violence in some locations while underemphasizing it in others. This can affect people's perceptions of their own locale, making them feel under siege and drawing more extreme repressive responses towards protesters (Body-

Gendrot 1995). The combination of exaggerated temporal and spatial clustering also produces a sense that the incidents are localized and not important to the national agenda.

Biased coverage of certain populations can have negative effects as well. The over-coverage of college student protest could lead to the impression that college kids were "going crazy" or that their behavior was more politically significant than the behavior of those partici-pating in street riots. In the end, this kind of bias in coverage could be producing or reinforcing a privilege hierarchy among protestors, with poor Blacks at the bottom and richer, white college students at the top (reinforcing the concerns Blacks expressed about the involvement of white college students in voter registration drives; e.g., McAdam 1988).

The errors of representation produced by reporters and editors become even more reified when governmental agencies study the problem and rely heavily on newspaper accounts (as did both the Kerner Commission and the 1967 U.S. Senate Subcommittee on Urban Violence). Not only can these agencies encourage certain perceptions about protests and protestors, they can also translate those perceptions into policy recommendations and action.

Distortion of Analyses

The bias that media selection produces is not limited to impressionistic or descriptive understandings of protest. These biases can also sharply distort even relatively simple numerical analyses as well. Consider one effect reported by Myers and Caniglia (2004): Newspapers over-reported riot events in U.S. cities that had relatively large Black populations. Over-reporting events in this way tends to over-estimate the effects of the size of the Black population on the incidence of rioting as we will illustrate with a simple simulation.

For example, suppose that there is no relationship between the Black population of a city and the number of riots that occur; and that there is a newspaper bias—similar to that docu-mented by Myers and Caniglia (2004)—that tends to ignore riots in cities that have higher percentages of whites in the population. We simulate this by drawing random numbers from 0-100 to represent percentage of the population that was Black, and random numbers from 0-10 to represent the number of riots that occurred. If we create data pairs using this method, there should be no correlation between the two variables, given that they are all random numbers. One such set of 100 pairs is plotted in figure 3 where each dot represents a city.
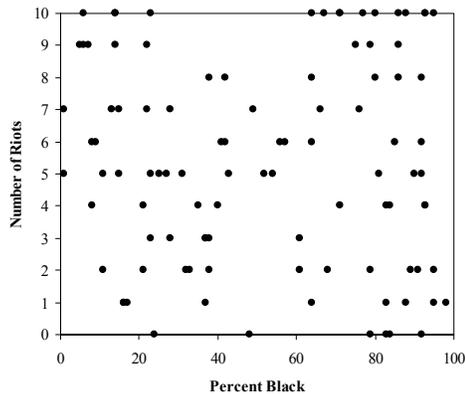
To represent bias, we adjust the hypothetical number of riots by reducing the number of riots observed in some cities. The higher the percentage of whites in the city, the more we reduce the number of riots.[3] This procedure effectively moves data points presented in figure 3 down the vertical axis (number of riots) toward zero—and does so even more for data points which are lower on the horizontal axis (cities where a low percentage of population is Black). This results in the data configuration presented in figure 4.

It is easy to see in this hypothetical example that the bias process will produce or increase the relationship between percent Black and the number of riots. We conducted 100 repetitions of this procedure, and although the original Pearson correlation between percent Black and the number of riots averaged approximately zero over all trials (.0049), once the bias procedure had been applied, the Pearson correlation increased to .32. We can see that the bias produced a false relationship in the data. This kind of bias, then, can be a very serious problem for the analysis and interpretation of newspaper data. The complexities with multivariate analyses are even more troublesome, difficult to predict, and manage, as Hug and Wisler (1998) have previously stated.

Can Analyses be Corrected? The Instability of Bias
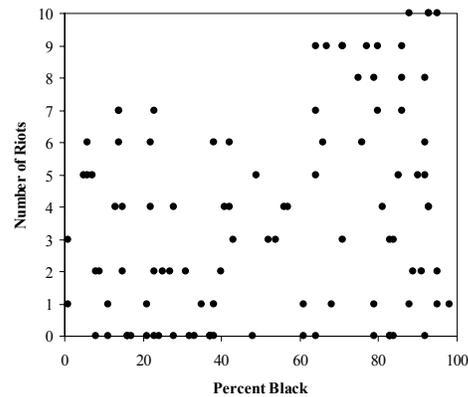
Because bias can have such serious effects on analyses, and because it is so ubiquitous in newspaper data, analysts have begun constructing procedures to correct for it. For the most part (e.g., Hug and Wisler 1998), these strategies have involved assessing selection effects in one limited time period and then using this information to correct for unknown bias in another,

Figure 3. Hypothetical Data Illustrating Random Relationship between Race and Number of Riots

Figure 4. Hypothetical Data Illustrating Biased Relationship between Race and Number of Riots

longer period. The challenge for these approaches is that they must assume that bias is consistent over time, both in terms of *selection factors* involved and the *amount* they affect selection. If not, the procedure may under-correct, over-correct, or possibly even increase the bias.

The consistency of bias over time is another issue about which scholars disagree. Some have concluded that bias is relatively stable, at least within a single newspaper, and therefore these time-shifted adjustments can be helpful (Olzak 1992; Franzosi 1987; McCarthy et al. 1996; Koopmans 1995; Barranco and Wisler 1999). Others have concluded that bias is inherently unstable and that corrective strategies relying on it cannot be trusted (Oliver and Myers 1999; Oliver and Maney 2000; Myers and Caniglia 2004). Unfortunately, we think that the research record supports the latter view, and that many factors contribute to marked instability in the newspaper record.

One paper that is often cited as evidence that bias is fairly consistent is McCarthy and colleagues' (1996) assessment of bias in the coverage of Washington, D.C. protest. In fact, McCarthy et al. (1996: 496) do conclude that "the national print media provides an amazingly stable portrait of the churning mixture of protest forms, purposes, and contexts in Washington D.C. during 1982 and 1991." However, the evidence presented in the article shows that the coverage provided by the *New York Times* and *Washington Post* varies to a greater extent than this conclusion suggests. If one reviews the selection factors examined across these two newspapers and over the two years of the study, the only consistently significant variable is demonstration size. In 1982, there is only one other consistently significant variable—the Middle-East Lebanese War, and in 1991, there is only one more—the Gulf War.[4] These differences mean that quite different selection processes were operating.

A similar mismatch exists between claims and evidence in Barranco and Wisler (1999: 316): "Apart from this [regional focus change], the bias structure of newspapers displays an astonishing stability over time and, therefore, provides a strong basis for valid observations of patterns of change in protest." However, in their analysis of stability over time in one national newspaper, the only significant variables are violence, large demonstrations, and squatter movements. And even these differ from the significant variables in the comparison of national and local newspapers, where squatter movements are no longer significant for the national newspaper. In sum, Barranco and Wisler's conclusion of relatively stable bias simply means that violence and size predict an increased likelihood of coverage of an event. Comparing this finding with the authors' additional findings—that the type of movements covered fluctuates, that areas of coverage may change, and that "patterns of change in protest" may reflect news selection factors or changing cultural contexts rather than changing patterns of protest—characterizing selection bias as "astonishingly" stable seems overstated.

Table 1. Logistic Regression Models Predicting Coverage of Riots in the *New York Times,* 1968-1969

	1968	(S.E.)	1969	(S.E.)
Distance				
Distance from NYC	-.000618	(.00052)	-.00149***	(.00041)
Distance from NYC^2	1.08×10^{-7}	(2.0×10^{-7})	3.05×10^{-7}*	(1.3×10^{-7})
Event in NYC	--[a]		--[a]	
Event in NY State	1.73***	(.38)	2.30***	(.45)
Event Intensity [b]				
Intensity Composite	.391	(.24)	.299	(.25)
Intensity × Distance	.00163*	(.00066)	.000680	(.00062)
Intensity × Distance^2	-4.76×10^{-7}	(2.8×10^{-7})	7.13×10^{-8}	(2.8×10^{-7})
Event Type				
Secondary School	-.328	(.26)	-.813**	(.27)
College or University	1.34***	(.37)	1.37***	(.31)
Street or Prison Event	--[c]		--[c]	
Proportion Black (Sqrt)	1.84**	(.60)	1.22	(.65)
Event Density	-.166	(.12)	-.513*	(.21)
Day of the Week [c]				
Sunday	.347	(.41)	.113	(.49)
Monday	.531	(.35)	.994*	(.40)
Tuesday	-.201	(.35)	.532	(.40)
Wednesday	-.118	(.40)	.565	(.40)
Thursday	.047	(.34)	.972*	(.39)
Saturday	-.0918	(.37)	.803	(.46)
Constant	-.683	(.65)	.615	(.79)
Model χ^2 (df)	139.54	(16)	201.42	(16)
N	535		579	
Pseudo R^2	.193		.269	

Notes: * $p < .05$, ** $p < .01$, *** $p < .001$ (two-tailed test).
[a] The New York City dummy variable is excluded because all NYC events in 1969 were covered and perfect prediction is not mathematically possible in logistic regression. We must simply report that all 1969 NYC events were covered and, in contrast, the NYC dummy was not significant when added to the 1968 model presented here.
[b] The intensity composite was constructed via a factor analysis of deaths, injuries, arrests, arson, and event duration. See Myers and Caniglia 2004 for details.
[c] Excluded dummy variable categories are Street/Prison events and Friday events.

Other analyses also demonstrate instability in bias across time and across newspapers (Oliver and Myers 1999; Oliver and Maney 2000). For instance, Myers and Canigla (2004) show that the percentage of events covered by a single newspaper varies markedly from month-to-month (from a high of over 60 percent to a low of under 20 percent). Beyond simply changing numbers of events, these data also demonstrate that the predictors of bias (selection factors) shift over time as well. In table 1, we reanalyze the bias analyses from Myers and Caniglia (2004), but this time we subdivide the data by year. As is immediately apparent, there is little consistency in the overall predictor model from year to year. In 1968, the selection model includes four significant variables and in 1969, it includes seven. Only two of these variables are the same (the college dummy variable and the NY State dummy variable). Thus, if we used the1968 model to correct the 1969 data, we would correct for *Proportion Black*, and thus erroneously adjust the model because that variable did not significantly effect selection in 1969. Likewise, the correction would not use the

distance factors to adjust estimates because they are not significant in the 1968 model, and would thereby ignore the necessary correction in 1969.

Beyond what has already been established empirically about the instability of bias, there are other reasons to believe that the selection mechanisms newspapers use change over time. Although these issues have not been empirically verified, they provide serious cause for concern— not only when applying corrections based on period sampling, but also when considering whether we are even collecting the same type of data from one period to the next. For example, one of the oft-touted advantages of newspaper data is that they provide a continuous record of events over a long period of time. But do they? For one thing, news holes expand and contract substantially over time. When an issue is hot and at the peak of its attention cycle (McCarthy et al. 1996), its news hole is relatively large. Newspaper editors are willing to devote more space to coverage of the issue and related events. At the same time, when there are extraordinary periods of protest activity, the news hole may not be elastic enough to accommodate the same proportion of events that it might in more peaceful times.

Second, beyond news-hole fluctuations, there are also major technological changes that fundamentally alter both the audience and the news gathering mechanisms for newspapers. An early critique of newspaper data (Danzger 1975) connected reports of collective behavior in a city with the presence of a news wire office in that city. These findings subsequently came under criticism (Snyder and Kelly 1977), but the importance of technologies like the wire service cannot be ignored. Without the wire service, newspapers would be severely handicapped in their attempts to report events occurring anywhere outside the local environment. Thus, the introduction of the wire service changes the news selection mechanism and a different set of events will likely appear in the newspaper as a result. Likewise, the use of the telephone, telegraph, fax, and now satellites, digital photography, the Internet, and web logs (blogs), all change the possibilities for gathering news and change the events reported and their timing.

Similarly, technology has wildly changed the distribution of newspapers. It is now possible to receive an on-time daily subscription to the *New York Times* anywhere in the United States. Thus, the audience has changed from when the *Times* was distributed more locally. This influences audience demand and changes what is reported in the *Times*, and thus what gets into our social science data. Changes in the numbers and types of events reported may, therefore, reflect changes in technology and journalistic practices, and not necessarily provide an accurate indication of change in the events themselves (Starobin 1995; Westerståhl and Johansson 1986).

PROSPECTS FOR MEDIA BASED DATA

The problems we have summarized above are serious, affect all newspaper data, and are not simple to correct. These problems are compounded by the wide use of newspaper data and the unquestioned acceptance of these data by many analysts and readers—an acceptance that has serious implications. McAdam, Sampson, Weffer, and MacIndoe (2005) recently argued that theory about protest and social movements has been unduly influenced by our overreliance on the U.S. civil rights movement as an empirical exemplar. By the same token, we would argue that our understandings of protest have also been overly influenced by our reliance on, and acceptance of, newspaper data—data that focuses on large, violent, regionally concentrated events, to the neglect of many other kinds of social movement activity.

Even those who have examined the newspaper-bias literature have a tendency to conclude that the data are not really that deficient—at least not worse than other social science data—and since there is no other choice, we should keep using what we have (Earl et al. 2004).

It is misleading to assume, however, that newspaper data are as good as other social science data. The biases in other kinds of social science data have been studied much more thoroughly and, because these biases are much better understood, correctives (at least partial) are commonly

used in analyses. For example, Earl et al. (2004: 77) write, "when considering selection bias, newspaper data compare favorably to bias in non-response in surveys." But in many surveys, we have the ability to model the patterns of missing data, have additional information collected from non-respondents, or have other information about the population. In newspaper data, even our most "complete" data misses an unknown amount of cases produced by highly unstable selection mechanisms. Therefore, nonresponses cannot, at this point, be reasonably modeled.

Furthermore, no matter how we improve our collection of newspaper data via more comprehensive data collection or better sampling, we will never be able to remove the fundamental source of bias—the practices of newspaper reporters and editors. Even if we somehow acquired the ability to record a complete census of every protest event reported in newspapers, the data would still be a flawed representation of protest due to the filters used to select them into the newspapers in the first place. Given these inherent difficulties and limitations, what can we do to improve newspaper data and the research that flows from them?

Acknowledging the Limitations of Newspaper Data

The first step in improving matters is, of course, admitting that there is a problem. Analysts need to take known biases seriously and take care to interpret their results in light of the potential problems. But, as is apparent from reading the literature, few do. Authors sometimes acknowledge in their papers that newspaper data are biased, but they do not carry this concern forward into their analysis or seem to think about the implications for their findings and interpretations.

To get a sense of how researchers treat the limitations of newspaper data and how they have reacted to the literature on media selection, we examined the five most recently published years of the *American Journal of Sociology* and the *American Sociological Review* (1999-2004). During that period, these journals published seventeen articles that used data collected from newspapers. Although this is a fairly small number of articles, and is obviously not representative of the literature in total, the placement of these pieces make them among the most influential of their type. They presumably have passed a demanding level of scrutiny, and as a result, provide a model of how to handle newspaper data for subsequent work. Among these articles, there were substantial differences in how newspaper data were used, and how findings were qualified because of potential bias or selection issues.

Many of these articles (ten) used newspaper data in ways that are not subject to the kinds of problems raised by selection bias research. For example, Barnett and Woywode (2004) examine the relationship between ideological orientations of Austrian newspapers and their failure or growth. Closely related, some studies examined the effects of media discourse on some other process (Koopmans and Olzak 2004; Koopmans and Statham 1999; McAdam and Su 2002)[5] or used media discourse as the dependent variable (Ferree 2003). When newspaper data are used in this manner—to represent what appears in the media rather than as a proxy for some underlying social phenomenon, many of the concerns related to selection bias are eliminated or, at least, markedly reduced.[6]

Another group of articles used newspaper data as a triangulated source in an attempt to qualitatively examine understandings and definitions of events (Vinitky-Seroussi 2000; Gerteis 2002; Isaac 2002; Carr 2003; Mueller 1999; Steinberg 2003). Given the purposes of the studies and how newspaper data were used in the context of other data, the selection issues seem somewhat less problematic. Nevertheless, authors need to take care that implied representativeness is carefully qualified. When one states that "if the data are roughly representative. . . , they tend to affirm several suppositions" (Steinberg 2003: 475), then the need to use the findings "cautiously" still exists (2003: 472).

The remaining studies (seven) are those that are of greatest concern in the present context. These use newspaper data to construct protest or movement event counts which are then used as outcome variables. As might be expected, there is a range of approaches to the relevant

methodological concerns. On one end, the methodological concerns can be completely ignored. In his examination of the role of political opportunity and threat in generating protest, Almeida (2003) relies primarily on one daily El Salvadoran newspaper and fails to discuss any methodological concerns regarding the use of newspapers as a data source.

On the other end, authors could point out the problems with newspaper data and actually engage these problems when they discuss their own findings. This occurred in only two of the studies we surveyed. In a study of Asian American collective events, Okamoto (2003) discusses potential biases, her strategies for decreasing the influence of the biases,[7] and consistently reminds the reader that these events are only the most "publicly visible collective events" (2003: 818). Oliver and Maney's (2000) study of how political processes affect both protest and protest coverage also contains a thorough critique of newspaper data as part of their conclusions.

Somewhere in the middle are those who provide what might be called lip service: they mention in their methods section that questions have been raised about the validity of newspaper data and affirm that caution is necessary during interpretation. When the time comes to interpret their data, however, the potential implications are never discussed. Four of the seven articles that used newspaper event counts followed this pattern and we have observed this pattern repeatedly in many journals over the past decade.

Often, these articles also misinterpret methodological findings from the media bias literature. For example, Koopmans and Olzak (2004) acknowledge bias problems, but claim these are not particularly relevant to their study because they are studying public discourse within the media. While the concerns are somewhat diminished in this context, they are not eliminated. More importantly, Koopmans and Olzak also perform an analysis of protest events using a variable developed from newspaper accounts. This variable is, as a result, subject to media selection bias.

Earl et al. (2003) admit that the literature shows that less intense events are less likely to be captured by newspaper data, but they dismiss this concern by stating that the time period they are studying is different than the time periods examined in bias studies.[8] Olzak and Uhrig (2001) fail to mention any bias concerns when they conclude that tactical overlap does not differ for violent and nonviolent protest events—even though the presence of violence has consistently been shown to influence the likelihood of coverage in newspapers (Snyder and Kelly 1977; Oliver and Myers 1999).

Distance-related bias is another factor that is often neglected. Myers and Caniglia (2004) demonstrated that non-New York state events are less likely to be covered in the *New York Times* than those inside New York. Consonant with this finding, Earl et al. (2003) limit their study only to New York State events as a strategy to reduce distance-related bias. However, Myers and Caniglia also found that after controlling for New York State events, there was still a substantial bias toward coverage of New York City events, and a decay of coverage over distance. Thus, data collected only for the State of New York is not fully representative of State events because there is an inordinate emphasis on New York City events relative to upstate ones. The findings, therefore, are heavily dependent on cultural and social processes in New York City—processes that are likely to be unique given the special social location of that city.

Sampling bias has also been consistently downplayed, particularly by those who have used the Prodat dataset.[9] The dataset was built using a strategy that examined Monday editions of the newspapers along with the entire week of papers for every fourth week. Olzak and Uhrig (2001) use the Monday sample based on Koopman's (1995) finding that the only systematic differences between the two components of the Prodat dataset is that the Monday subsample underestimates the number of events, and Rucht's (1998) finding that labor/union sector events were more likely to be undercounted using the Monday-only method. Rucht (1998) *does not* claim, however, that new social movement events *are not* undercounted. Rather, he merely says that they are less likely to be undercounted relative to labor/union

sector events. More importantly, Koopmans's and Rucht's comparisons of the two samples are just that: comparisons between two biased samples (just as all selection studies are). Thus we cannot use this comparison to establish a lack of bias. Rather, we only know that one sample contains more or less of certain types of events than the other.

Finally, researchers have consistently been tempted to argue that even if newspaper coverage is biased, that these biases are stable across time within the same newspaper (Jenkins, Jacobs, and Agnone 2003). If this were true, certain kinds of analyses would make more sense, but, as we have discussed above, the evidence against consistency has been moun-ting, and the claim of consistent bias has become implausible.

Steps Toward Improvement

The treatment of media bias in recent literature suggests that researchers are not yet taking selection bias seriously. After reviewing the literature and working with newspaper data ourselves, both in substantive and methodological studies, we think it is clear that these data face difficult challenges. To improve the use of newspaper data, researchers need to think more carefully about if they are even appropriate for the task at hand. Not every study we would like to do is actually viable, and one should consider carefully whether or not the data can adequately respond to the challenges of the analysis. If scholars do proceed with a study, they should think and write about the limitations of their data, the biases that likely exist in them, and inform the reader straightforwardly about how these biases may be affecting the results of their study. These, we propose, are the minimal steps necessary to proceed with newspaper data given the small mountain of evidence about bias that now exists.

Another important step is to collect better data. Unfortunately, improving the data is easier said than done. Gathering data from additional media sources and from non-media sources (such as government records) is expensive, time consuming, and can produce problems with comparability. In addition, more data are not always better data. As Myers and Caniglia (2004) illustrated, adding just a small number of sources could have the effect of increasing some of the discrepancies already present in newspaper data. The development of electronic media databases may provide new ways of accessing protest data less expensively, but at present, even fairly comprehensive resources like Lexis-Nexis have unstable content and important gaps in their scope. Even if these resources do become a comprehensive clearinghouse for media records, they still cannot escape the bias that selects events into the media record in the first place.

Moreover, we must continue to study the problem and shift some of our methodological research to focus on solutions. For example, one notion that has not yet been investigated, but might bear fruit is the parallels and discontinuities between newspaper bias and survey non-response bias. The idea that events not reported in newspapers can be thought of as non-responders in surveys is certainly an interesting notion, and although there is not yet any evidence showing that this is the case, a comparison of these two missing data processes may lead to some additional insight into the management of newspaper data bias.

We must also bear in mind that some types of research using newspaper data are not as problematic as others. In this article, we identified several studies in which the use of newspaper data was intended to represent media content, rather than an estimate of real world events. Under these circumstances, media data are much closer to what they purport to represent and the bias issues are markedly reduced. Research that uses event counts as a dependent variable faces more difficult problems.

Finally, our critique should not be construed as meaning that newspaper data cannot or should not be used for research purposes. We think that the challenges for newspaper data are formidable and are going to require much more work, but our main intention is to issue a series of pleas: first, that scholars take selection bias findings more seriously; second, that analysts attend to the possible effects of bias in their studies, and write about the limitations of the data and how they might be affecting their analyses; third, that scholars continue to conduct research

on newspapers as a data source; and last, that we all avoid the temptation to use newspaper data where inappropriate. These are small but necessary steps for a qualitative improvement in the use of newspaper data in event research.

NOTES

[1] For more on corporate interest and the distortion of the objectivity in the media, see Bozell and Baker 1990; Novick 1988; Schudson 1978; and Shumway 1996.

[2] It is possible to include additional effects in the model (corporate interests influencing audience demands, for example), but we chose this more parsimonious model to provide a heuristic guide to the most important media selection effects.

[3] The formula used to reduce the number of riots was simply (100 - percent white)/25, which reduced the number of riots anywhere from 0 to 4.

[4] In McCarthy et al.'s full model represented in table 5 (p. 491), four variables change sign and/or significance between 1982 and 1991. Negative and significant coefficients are obtained for vigil/picket, campaign, march and rally, and rally for 1982. However, vigil/picket and march and rally become positive and non-significant in 1991, while campaign and rally become negative and non-significant. In the full model for 1982 in table 6, counter demonstration is negative and non-significant for *NYT*, and negative and significant for WP. Vigil/picket is negative and non-significant for *NYT*, and negative and significant for WP. The full model for 1991 alone, table 7, shows even more inconsistencies. In the *NYT,* weekend is negative and significant, march and rally is positive and non-significant, labor is positive and significant, and healthcare is negative and non-significant. In contrast, for the WP, weekend is non-significant, march and rally is significant, labor is non-significant, and health care is positive and significant. When we compare tables 5, 6 and 7, more evidence is found of instability in the data results. Although weekend was negative and significant in their full model for 1982 and 1991, in the separate tables for 1982 and 1991, weekend is only negative and significant for the *NYT* in 1991 (table 7). In table 5, campaign, march/rally, vigil/picket, and rally are negative and significant for 1982. In table 6 (1982 alone), for both *NYT* and WP march/rally, and rally are negative and non-significant. Campaign is positive and non-significant in NYT, and negative and non-significant in WP. Vigil/picket is negative and non-significant in *NYT*. Anti-nuclear and ERA are positive and significant in Table 5 for 1982. However, in table 6, antinuclear becomes negative and non-significant in *NYT* and positive and non-significant in WP; ERA becomes non-significant for both news sources.

[5] Koopmans and Olzak (2004) use newspaper data both as an explanatory variable for media discourse and as an outcome variable for protest event counts. The use of these data for event counts is discussed below.

[6] There remains the issue that the newspaper data examined by the researchers may be a biased subset of all media data (ignoring, for example, other media sources such as the television and Internet, and other newspapers which might contribute something rather different to the media discourse), a problem that these articles do not address.

[7] She avoids use of regional and local news media, as well as ethnic and panethnic newspapers as they all tend to have a more regional focus. She utilizes more than one newspaper, avoids sampling strategies that focus only on issues from one or two days, and utilizes a "generic descriptor" in an attempt to capture smaller events, indoor events and institutional events.

[8] Earl et al. (2004) also use the different time period argument to minimize bias concerns about the likelihood of coverage.

[9] Prodat is a database that has been developed by Friedhelm Neidhardt and Dieter Rucht at the Berlin Center of Science and Social Research which utilizes reports of protest events from two newspapers with wide circulation.

REFERENCES

Almeida, Paul D. 2003. "Opportunity Organizations and Threat-Induced Contention: Protest Waves in Authoritarian Settings." *American Journal of Sociology* 109(2): 344-400

Bagdikian, Ben H. 2000. *The Media Monopoly*. Boston, MA: Beacon.

Banks, Arthur S. 1997. *Cross-National Times Series Data Archive*. Binghamton, NY: Computer Solutions Unlimited.

Barnett, William P., and Michael Woywode. 2004. "From Red Vienna to the Anschluss: Ideological Competition among Viennese Newspapers During the Rise of National Socialism." *American Journal of Sociology* 109(6): 1452-1499.

Barranco, José, and Dominique Wisler. 1999. "Validity and Systematicity of Newspaper Data in Event Analysis." *European Sociological Review* 15(3): 301-322.

Basil, Michael. 1996. "Identification as a Mediator of Celebrity Effects." *Journal of Broadcasting and Electronic Media* 40(4): 478-495.

Body-Gendrot, Sophie. 1995. "Urban Violence: A Quest for Meaning." *New Community* 21: 525-536.

Bond, Doug, J. Craig Jenkins, Charles L. Taylor, and Kurt Schock, 1997. "Maping Mass Political Conflict and Civil Society." *Journal of Conflict Resolution* 41(4): 553-579.

Bozell, L. Brent and Brent H. Baker, eds. 1990. *And That's the Way It Isn't: A Reference Guide to Media Bias*. Alexandria, VA: Media Research Center.

Burch, Philip H. 1972. *The Managerial Revolution Reassessed*. Lexington, MA: Heath.

Carr, Patrick J. 2003, "The New Parochialism: The Implications of the Beltway Case for Arguments Concerning Informal Social Control." *American Journal of Sociology* 108(6): 1249-1291.

Chen, Ning. 2003. "A Study of the Mechanism of Psychological Processing of Celebrity Advertisement Spokesmen of Different Ages." *Psychological Science* 26: 37-40.

Cooper, Purvis, Michael Burgoon, and Debra L. Roter. 2001. "An Expectancy-value Analysis of Viewer Interest in Television Prevention News Stories." *Health Communication* 13: 227-240.

Danzger, M. Herbert. 1975. "Validating Conflict Data." *American Sociological Review* 40(5): 570-584.

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.

Earl, Jennifer, Sarah A. Soule, and John D. McCarthy. 2003. "Protests Under Fire? Explaining Protest Policing." *American Sociological Review* 68(4): 581-606.

Feree, Myra Marx. 2003. "Resonance and Radicalism: Feminist Framing in the Abortion Debates of the United States and Germany." *American Journal of Sociology* 109(2): 304-344.

Fiske, Susan T. 1980. "Attention and Weight in Person Perception: The Impact of Negative and Extreme Behavior." *Journal of Personality and Social Psychology* 38: 889-906.

Franzosi, Roberto. 1987. "The Press as a Source of Socio-Historical Data: Issues in the Methodology of Data Collection from Newspapers." *Historical Methods* 20(1): 5-16.

———. 2004. *From Words to Numbers: Narrative, Data and Social Science*. New York: Cambridge University Press.

Galambos, Louis. 1975. *Public Image about Big Business, 1880-1940: A Quantitative Study in Social Change*. Baltimore, MD: Johns Hopkins.

Gamson, William A., and Gadi Wolfsfeld. 1993. "Movements and Media as Interacting Systems." *Annals of the American Academy of Political and Social Science* 528(July): 114-125.

Gamson, William A., David Croteau, William Hoynes, and Theodore Sasson. 1992. "Media Images and the Social Construction of Reality." *Annual Review of Sociology* 18: 373-393.

Gans, Herbert J. 1979/1980. *Deciding What's News: A Study of CBS Evening News, NBC Nightly News, Newsweek and Time*. New York: Vintage.

Gerteis, Joseph. 2002. "The Possession of Civic Virtue: Movement Narratives of Race and Class in the Knights of Labor." *American Journal of Sociology* 108(3): 580-615.

Ghanem, Salma. 1996. *Media Coverage of Crime and Public Opinion: An Exploration of the Second Level of Agenda Setting*. Ph.D. dissertation, University of Texas at Austin.

Gitlin, Todd. 1980. *The Whole World is Watching: Mass Media in the Making and Unmaking of the New Left*. Berkeley: University of California Press.

Goren, Dina. 1980. "The News and Foreign Policy: An Examination of the Impact of the News Media on the Making of Foreign Policy." *Research in Social Movements, Conflicts and Change* 3: 119-41.

Gurr, Ted R. 1968. "A Causal Model of Civil Strife: A Comparative Analysis Using New Indices." *American Political Science Review* 62: 1104-1124.

Harrison, Kathryn. 2001. "Too Close to Home: Dioxin Contamination of Breast Milk and the Political Agenda." *Policy Sciences* 34(1): 35-62.

Hazelwood, Leo A., and Gerald T. West. 1974. "Bivariate Associations, Factor Structures, and Substantive Impact: The Source Problem Revisited." *International Studies Quarterly* 18: 317-337.

Herman. Edward S. 1995. *Triumph of the Market: Essays on Economics, Politics, and the Media*. Boston: South End Press.

Herman, Edward S., and Noam Chomsky. 1988. *Manufacturing Consent*. New York: Pantheon Books.

Hocke, Peter. 1998. "Determining Selection Bias in Local and National Newspaper Reports on Protest Events." Pp. 131-163 in *Acts of Dissent: New Developments in the Study of Protest,* Dieter Rucht, Ruud Koopmans, and Friedhelm Neidhardt, eds. Berlin: Wissenschaftszentrum Berlin Fur Sozialforschung.

Hug, Simon, and Dominique Wisler. 1998. "Correcting for Selection Bias in Social Movement Research." *Mobilization: An International Journal* 15(3): 141-161.

Isaac, Larry. 2002. "To Counter 'the Very Devil' and More: the Making of Independent Capitalist Militia in the Gilded Age." *American Journal of Sociology* 108(2): 353-405.

Jackman, Robert W., and William A. Boyd. 1979. "Multiple Sources in the Collection of Data on Political Conflict." *American Journal of Political Science* 23: 434-458.

Jeffries, Vincent, Ralph H. Turner, and Richard T. Morris. 1971. "The Public Perception of the Watts Riot as Social Protest." *American Sociological Review* 36(3): 443-451.

Jenkins, J. Craig, and Charles Perrow. 1977. "Insurgency of the Powerless: Farm Worker Movements (1946-1972)." *American Sociological Review* 42(2): 249-268.

Jenkins, J. Craig, and Craig M. Eckert. 1986. "Channeling Black Insurgency: Elite Patronage and Professional Social Movement Organizations in the Development of the Black Movement." *American Sociological Review* 51(6): 812-829.

Jenkins, J. Craig, David Jacobs and Jon Agnone. 2003. "Political Opportunities and African-American Protest, 1948-1997." *American Journal of Sociology* 109(2): 277-303.

Kiousis, Spiro. 2004. "Explicating Media Salience: A Factor Analysis of *New York Times* Issue Coverage During the 2000 U.S. Presidential Election." *Journal of Communication* 54: 71-87.

Kiousis, Spiro, Philemon Bantimaroudis, and Hyun Ban. 1999. "Candidate Image Attributes: Experiments on the Substantive Dimension of Second Level Agenda Setting." *Communication Research* 26: 414-428.

Klatzsky, Roberta L. 1975. *Human Memory: Structures and Processes*. New York: Freeman.

Knobloch, Silvia, Matthias Hastall, Dolf Zillmann, and Coy Callison. 2003. "Imagery Effects on the Selective Reading of Internet Newsmagazines." *Communication Research* 30: 3-29.

Knobloch, Silvia, Dolf Zillman, Rhonda Gibson, and James A. Karrh. 2002. "Effects of Salient News Items on Information Acquisition and Issue Perception." *Zeitschrift fur Medienpsychologie*, 14: 14-22.

Koopmans, Ruud. 1995. "The Use of Protest Data in Comparative Research: Cross-National Comparability, Sampling Methods and Robustness." Pp. 90-110 in *Acts of Dissent: New Developments in the Study of Protest,* Dieter Rucht, Ruud Koopmans, and Friedhelm Neidhardt, eds. Berlin: Wissenschaftszentrum Berlin Fur Sozialforschung.

Koopmans Ruud, and Susan Olzak. 2004. "Discursive Opportunities and the Evolution of Right-wing Violence in Germany." *The American Journal of Sociology* 110(1): 198-230.

Koopmans, Ruud, and Dieter Rucht. 2002. "Protest Event Analysis." Pp. 231-259 in *Methods of Social Movement Research*, Bert Klandermans and Suzanne Staggenborg, eds. Minneapolis: University of Minnesota Press.

Koopmans, Ruud, and P. Statham. 1999. "Challenging the Liberal Nation-State? Postnationalism, Multiculturalism, and the Collective Claims Making of Migrants and Ethnic Minorities in Britain and Germany." *The American Journal of Sociology* 105(3): 652-696.

Kriesi, Hanspeter, Ruud Koopmans, Jan Willem Duyvendak, and Marco G. Guigni. 1995. *New Social Movements in Western Europe*. Minneapolis: University of Minnesota Press.

Lee, Martin A., and Norman Solomon 1990. *Unreliable Sources*. New York: Carol Publishing Group.

Leyens, J. P., and V. Y. Yzerbyt. 1992. "The Ingroup Overexclusion Effect: Impact of Valence and Confirmation on Stereotypical Information Search." *European Journal of Social Psychology* 22: 549-570.

Lieberson, Stanley, and Arnold R. Silverman. 1965. "The Precipitants and Underlying Conditions of Race Riots." *American Sociological Review* 30: 887-898.

Lipsky, Michael, and David J. Olson. 1976. "The Processing of Racial Crisis in America." *Politics and Society* 6: 478-499.

MacGregor, Brent. 1997. *Live, Direct, and Biased? Making Television News in the Satellite Age*. New York: Arnold Publishers.

Maney, Gregory M., and Pamela E. Oliver. 2001. "Finding Event Records: Timing, Searching, Sources." *Sociological Methods and Research* 29: 131-169.

McAdam, Doug. 1982. *Political Process and the Development of Black Insurgency, 1930-1970*. Chicago: University of Chicago Press.

———. 1988. *Freedom Summer*. New York: Oxford University Press.

McAdam, Doug, and Yang Su. 2002. "The War at Home: Antiwar Protests and Congressional Voting, 1965 to 1973." *American Sociological Review* 67(5): 696-725.

McAdam, Doug, Robert J. Sampson, Simon Weffer, and Heather MacIndoe. 2005. "'There will be Fighting in the Streets': The Distorting Lens of Social Movement Theory." *Mobilization: An International Journal* 10(1): 1-18.

McCarthy, John D., Clark McPhail, and Jackie Smith. 1996. "Images of Protest: Dimensions of Selection Bias in Media Coverage of Washington Demonstrations, 1982 and 1991." *American Sociological Review* 61: 478-499.

McCombs, Maxwell, Juan Pablo Llamas, Esteban Lopez-Escobar, and Federico Rey. 1997. "Candidate Images in Spanish Elections: Second-level Agenda-Setting Effects." *Journalism and Mass Communication Quarterly* 74: 703-717.

McCombs, Maxwell, and Donald Shaw. 1972. "The Agenda-setting Function of Mass Media." *Public Opinion Quarterly* 36: 176-187.

McCombs, Maxwell, and Salma Ghanem. 2001. "The Convergence of Agenda Setting and Framing," Pp. 67-83 in *Framing Public Life: Perspectives on Media and Our Understanding of the Social World*, Stephen D. Reese, Oscar H. Grandy, Jr., and August E. Grant, eds. Mahwah, NJ: Erlbaum.

Meyer, David S. 1993. "Protest Cycles and Political Process: American Peace Movements in the Nuclear Age." *Political Research Quarterly* 46: 451-479

———. 1995. "Framing National Security: Elite Public Discourse on Nuclear Weapons During the Cold War." *Political Communication* 12: 173-192.

Monroy, Tiffany, Dyan Kuniko, and Daniel J. Myers. 2004. "Fanning the Flames? Riot Commissions and the Mass Media." *Race, Gender & Class* 11(1): 136-157.

Mueller, Carol. 1999. "Escape from the GDR, 1961-1989: Hybrid Exit Repertoires in a Disintegrating Leninist Regime." *The American Journal of Sociology* 105(3): 697-758.

———. 1997a. "International Press Coverage of East German Protest Events, 1989." *American Sociological Review* 62: 820-832.

———. 1997b. "Media Measurement Models of Protest Event Data." *Mobilization: An International Journal* 2: 165-184.

Myers, Daniel J. 1997. "Racial Rioting in the 1960s: An Event History Analysis of Local Conditions." *American Sociological Review* 62: 94-112.

———. 2000. "The Diffusion of Collective Violence: Infectiousness, Susceptibility, and Mass Media Networks." *American Journal of Sociology* 106: 173-208.

Myers, Daniel J., and Beth S. Caniglia. 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.

National Advisory Commission on Civil Disorders. 1968. *Report of the National Advisory Commission on Civil Disorders*. New York: Bantam Books.

Novick, Peter. 1988. *That Noble Dream: The "Objectivity Question" and American Historical Profession*. Cambridge, MA: Cambridge University Press.

Okamoto, Dina G. 2003. "Toward a Theory of Panethnicity: Explaining Asian American Collective Action." *American Sociological Review* 68(6): 811-842.

Oliver, Pamela E., and Gregory Maney. 2000. "Political Processes and Local Newspaper Coverage of Protest Events: From Selection Bias to Triadic Interactions." *American Journal of Sociology* 106(2): 463-505.

Oliver, Pamela E., and Daniel J. Myers. 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, Susan. 1982. "Ethnic Mobilization in Quebec." *Ethnic and Racial Studies* 5: 253-275.

———. 1992. *The Dynamics of Ethnic Competition and Conflict*. Stanford, CA Stanford University Press.

Olzak, Susan, and Johan L. Olivier. 1994. "The Dynamics of Ethnic Collective Action in South Africa and the United States: A Comparative Study." Paper presented at the International Sociological Association meetings, Bielefeld, Germany.

Olzak, Susan, and S. C. Noah Uhrig 2001. "The Ecology of Tactical Overlap." *American Sociological Review* 66(5): 694-717.

Paige, Jeffery. 1975. *Agrarian Revoultion: Social Movements and Export Agriculture in the Underdeveloped World*. New York: Free Press.

Parenti, Michael. 1993. *Inventing Reality: The Politics of the News Media*. New York: St. Martin's Press.

Perse, Elizabeth M. 1992. "Predicting Attention to Local Television News: Need for Cognition and Motives for Viewing." *Communication Reports* 5: 40-49.

Pratto, F., and Oliver, J. 1991. "Automatic Vigilance: The Attention Grabbing Power of Negative Social Information." *Journal of Personality and Social Psychology* 61: 380-391.

Richards, Assata, and John McCarthy. 2002. "The Role of Controversy and Celebrity in the Newspaper Coverage of the Million Man March and the Promise Keepers Stand in the Gap Rally." Manuscript, Pennsylvania State University.

Roberts, Marilyn, Wayne Wanta, and Tzong-Horng Dzwo. 2002. "Agenda Setting and Issue Salience Online." *Communication Research* 29: 452-465.

Rucht, Dieter, and Thomas Olemacher. 1992. "Protest Event Data: Collection, Uses and Perspectives." Pp. 76-106 in *Studying Collective Action*, Mario Diani and Ron Eyerman, eds. London: Sage.

———. 1998. "The Structure and Culture of Collective Protests in Germany since 1950." Pp. 30-54 in *The Social Movement Society*, David S. Meyer and Sidney Tarrow, eds. New York: Cambridge University Press.

Rucht, Dieter, Ruud Koopmans, and Friedhelm Neidhardt, eds. 1995. *Acts of Dissent: New Developments in the Study of Protest*. Berlin: Wissenschaftszentrum Berlin Fur Sozialforschung.

Rummel, Randolph J. 1966. "Dimensions of Conflict Behavior within Nations, 1946-59." *Journal of Conflict Resolution* 10: 65-73.

Ryan, Charlotte. 1991. *Prime Time Activism*. Boston: South End Press.

Schrodt, Philip A. 1994. "The Statistical Characteristics of Event Data." *International Interactions* 20: 35-53.

Schudson, Michael. 1978. *Discovering News: A Social History of American Newspapers*. New York, NY: Basic Books.

Shaw, Donald L., and Maxwell E. McCombs, eds. 1977. *The Emergence of American Political Issues: The Agenda Setting Function of the Press*. St. Paul, MN: West.

Shoemaker, Pamela J. 1988. "The Perceived Legitimacy of Deviant Political Groups: Two Experiments on Media Effects." *Communication Research* 9: 249-286.

Shorter, Edward, and Charles Tilly. 1974. *Strikes in France*. New York: Cambridge University Press.

Shumway, David. 1996. "Objectivity, Bias, Censorship." Pp 239-254 in *Making and Selling Culture,* R. Ohmann, ed. Hanover, NH: The University Press of New England.

Smith, Jackie, John D. McCarthy, Clark McPhail, and B. Augustyn. 2001. "From Protest to Agenda Building: Description Bias in Media Coverage of Protest Events in Washington, D.C." *Social Forces* 79 (4): 1397-1423.

Snyder, David, and William R. Kelly. 1977. "Conflict Intensity, Media Sensitivity and the Validity of Newspaper Data." *American Sociological Review* 42: 105-123.

Spilerman, Seymour. 1970. "The Causes of Racial Disturbances: A Comparison of Alternative Explanations." *American Sociological Review* 35: 627-49.

———. 1976. "Structural Characteristics of Cities and the Severity of Racial Disorders." *American Sociological Review* 41: 771-93.

Starobin, Paul. 1995. "A Generation of Vipers: Journalists and the New Cynicism." *Columbia Journalism Review* March/April: 25-32.

Steinberg. 2003. "Capitalist Development, the Labor Process, and the Law." *American Journal of Sociology* 109(2): 445-495.

Taylor, Charles Lewis, and Michael C. Hudson. 1972. *World Handbook of Political and Social Indicators*, 2nd ed. New Haven: Yale University Press.

Taylor, Charles Lewis, and David A. Jodice. 1983. *World Handbook of Political and Social Indicators*, 3rd ed. New Haven: Yale University Press.

Tilly, Charles. 1997. "Parliamentarization of Popular Contention in Great Britain, 1758-1834." *Theory and Society* 26: 245-273.

Vinitky-Seroussi, Vered. 2002. "Commemorating a Difficult Past: Yitzhak Rabin's Memorials" *American Sociological Review* 67(1): 30-51.

Wagers, Robert. 1992. "Dowquest and Dow Jones Text-Search: Which Works Best and When?" *Online* 16 (November):35-36, 38, 40-41.

Westerståhl, Jörgen, and Folke Johansson. 1986. "News Ideologies as Moulders of Domestic News." *European Journal of Communication* 1: 133-150.

Woolley, John T. 2000. "Using Media-Based Data in Studies of Politics." *American Journal of Political Science* 44: 156-173.