STUDYING ONLINE ACTIVISM: THE EFFECTS OF SAMPLING DESIGN ON FINDINGS*

Jennifer Earl†

Social movement scholars are increasingly interested in Internet activism but have struggled to find robust methods for identifying cases, particularly representative samples of online protest content, given that no population list exists. This article reviews early approaches to this problem, focusing on three recent case sampling designs that attempt to address this problem. The first approach purposively samples from an organizationally based sampling frame. The second approach randomly samples from a SMO-based sampling frame. The third approach mimics user routines to identify populations of "reachable" websites on a given topic, which are then randomly sampled. For each approach, I examine the sampling frame and sampling method to understand how cases were selected, outline the assumptions built into the overall sampling design, and discuss an exemplary research project employing each design. Comparisons of findings from these exemplar studies indicate that sampling designs are extremely consequential. I close by recommending best practices.

Information communication technologies (ICTs) are increasingly playing an important role in protest and broader social movements. It is critical, therefore, that scholars design and execute rigorous research programs to investigate how ICTs are used by protesters and organizers and how that usage affects social movements substantively and theoretically. This kind of investtigation often quickly turns to studying protest-relevant material that can be found online in search of both descriptive and causal insights. For instance, descriptive questions such as the percentage of protest-related websites that support offline protests and/or offer online avenues for protest participation are important issues, as are the relationships between organizational sponsorship and kinds of activities offered.

The standard resolution to the need for population estimates is to identify an excellent sampling frame and then randomly sample from it. Unfortunately, it has proven quite difficult to identify population lists of online protest content or actions that can serve as good sampling frames, which has substantially complicated case selection and made understanding the contours of protest content online much more difficult. In the face of such a daunting methodological dilemma, a variety of approaches to case selection have been pursued. For instance, case studies where cases were selected because of their notoriety, popularity, success, or the importance of the offline organizations sponsoring selected websites have been common (e.g., Bennett and Fielding 1999; Martinez-Torres 2001) because they side-step the need for sampling.

Nonetheless, a number of scholars have tried to move beyond single case studies to look at sets of websites, or what one could think of loosely as various types of samples. This article

---

* I would like to thank the National Science Foundation for generous support of this research through an NSF CAREER Award (SES-0547990) and to thank the many research assistants who worked on this project (see http://jearl.faculty.arizona.edu/internet_and_activism_lab for a full listing). Thanks also to Heidi Reynolds-Stenson and Misty Ring for their research assistance and to the anonymous reviewers and the special editor, Neal Caren, for their comments on previous drafts.

† Jennifer Earl is a Professor in the School of Sociology at the University of Arizona. Please direct correspondence to jenniferearl@email.arizona.edu.

© 2013 Mobilization: An International Quarterly 18(4): 389-406

examines three sampling designs in detail, each of which represents a distinct approach to constructing sampling frames (i.e., identifying cases at risk of being studied) and sampling methods (i.e., how cases from the sampling frame are drawn). In the first approach, larger sets of websites are constructed using knowledge of the movement, meaning that the sampling frame includes groups already known to the investigators. The sampling method is purposive. In the second approach, offline organizations listed in a well-known organizational directory are randomly sampled and then their websites are studied. Therefore, the sampling frame is the organizational directory and the sampling method is random. A final approach builds a sampling frame from search engine results and then randomly samples from those results. In this article, I describe the three approaches, unpack assumptions related to the overall sampling design that may impact substantive findings, and compare the three methods against one another using research from three representative projects.[1]

This comparison reveals that sampling design likely plays a large role in shaping the results of studies, significantly calling into question the generalizability of findings from many research projects. Specifically, findings show that: (1) sampling frames that rely on organizations as anchoring social movement activity produce notably partial views of the larger field of movement activity;[2] (2) organizationally focused sampling frames, whether the sampling method is random or purposive, underestimate the amount of online forms of activism but overestimate the amount of online facilitation of offline protest, when compared to samples from frames that include both organizationally produced and non-SMO-produced sites; (3) sampling frames and sampling methods affect estimated frequencies for specific actions; and (4) sampling design does not markedly influence estimates of very pervasive types of activities such as sharing news and information about a cause. Using these results, I close by suggesting best practices for future research.

PRIOR RESEARCH PRACTICES IN STUDYING "INTERNET ACTIVISM"

Although ICTs in protest are endemic today and research on the topic is increasingly prevalent, early research progressed more slowly and tended to focus on a small number of cases. For instance, the use of the Web by the Zapatistas and the global social justice movement received substantial early scholarly attention (Ayres 1999; Garrido and Halavais 2003; Martinez-Torres 2001; Smith and Smyth 2001; Van Aelst and Walgrave 2004; Wray 1999). Even where scholars examined other protest campaigns, they often pursued a single case study or studied, at most, a few cases together (e.g., Carty 2002; Earl and Schussman 2003, 2004; Gurak 1999, 1997; Gurak and Logie 2003; Schussman and Earl 2004).

Although case studies can be tremendously useful for addressing a variety of theoretical issues, a major shortcoming of case studies is that their aggregation cannot provide a high-quality image of larger population characteristics and/or dynamics. In other words, while incredibly helpful, case studies are able to contribute the most to fields in which specific cases can already be situated within larger populations of cases so that the generalizability of findings is easier to discern. One would suspect, then, that as case studies accumulated, there would be a move to examine larger populations of online activism.

Unfortunately, a significant hurdle stood in the way: there is no population list of activist websites or online activist activities that can be used as a sampling frame. Moreover, "work arounds" that have allowed survey centers to overcome the lack of a population list of households—such as random digit dialing—cannot be readily adapted to the online environment. When random digit dialing was created, it benefited from two key facts: (1) telephones were so widely adopted that most households could be assumed to be reachable by phone (i.e., telephone numbers could serve as the sampling frame); and (2) the patterning of phone numbers and area codes allowed random potential numbers to be generated for specific geographic areas (i.e., one did not even need to have a full list of elements from the frame to

randomly sample from it). In other words, the genius behind random digit dialing was that one did not need a population list of households or individuals so long as a random way to reach those individuals by phone could be created (i.e., random phone numbers). Unfortunately for students of online protest, there is no online analog to a phone number that can allow the random generation of meaningful Web addresses.[3]

Without a population list, or a clear-cut work around, researchers struggled to broaden the set of cases examined. One early approach that has turned out to be more popular in other areas of Web research than in research on online protest has been to create "communities" of websites that are connected to one another via hyperlinks (Thelwall 2004). In this kind of study, scholars choose a small to moderate number of "seed sites"—or starting locations—and then find all websites that are hyperlinked from those initial sites. One could imagine that this is similar to an extremely extensive snowball sample; much like including informants related to your current subject, this method allows researchers to capture all of the first-degree connections that a given site has to other websites. In fact, if scholars desired, they could also identify sites that were two or more links away from the seed site (i.e., site A links to site B, which links to site C).

Tools for this kind of research proliferated (e.g., Socibot and Issue Crawler to name two of the most popular), but several lingering problems have limited its application to social movement studies. First, websites link to other websites for a myriad of reasons that are poorly understood, implying that it is hard to know how to interpret a link and how to interpret larger networks of sites based upon hyperlinks. Are these networks actually structured communities? Are ties represented through hyperlinks very meaningful and/or comparable across sites? How often do links change, get updated, or get removed? Might there be poorly connected parts of larger communities or populations that would be hard to reach this way, such as would occur if sites and their allies intentionally did not link to another set of sites because of competition or animosity? This style of study has proven more useful in instances where the meanings of links are clear, but this method has thus far proven less useful for students of protest (Ackland 2009).

Second, from a methodological standpoint, one should not mistake what is essentially a snowball sample (often built on uncertain ties) as a definitive population for an area. For instance, if one began with five to ten global social justice websites, it would be hard to claim that links from those sites to other sites constitute the universe of global social justice material online. The true universe is certainly much wider and includes material not directly linked from the starting seed sites. Even if one extends this to two or three steps away from the original seed sites, making the network of sites expand substantially, it still would not necessarily encompass anything approaching a population of global social justice websites.

Third, even if one were not concerned about what links really mean, and hence what set of sites is being identified by the technique, only going two links away from the original sites could net tens of thousands to millions of pages, many of which are actually unrelated. For instance, Garrido and Halavais (2003) showed that by going two links away from their original seed sites, it is necessary to crawl a hundred thousand pages, which contain several million links. Barring major advances in areas such as machine learning, processing that amount of information in anything other than a cursory fashion is very difficult. Until such major advances occur, there is no computationally efficient way to discern which sites are meaningfully connected and which are not.[4]

I should note that there are likely areas of research where this method will gain more traction, even within social movement studies. For instance, network-based crawls using links between users are likely to be helpful for scholars studying social networking sites, such as Facebook, or other sites built around stylized social connections, such as Twitter. In such cases, one can safely make far more assumptions about the nature and social meaning of ties and can make reasonable assumptions about nodes being connected if one crawls far enough

from original seed users. But, those same assumptions are not as readily made for broader kinds of content and less well-defined linking practices.

With this approach yielding few returns for protest scholars interested in online content, researchers turned to sampling-based approaches, but with different simplifying assumptions designed to make the lack of a true population list tractable. In the next three sections, I outline these approaches in turn and introduce a research project that could be treated as an exemplar of each method.

PURPOSIVE ORGANIZATIONAL SAMPLES

One way to create a sampling frame when one lacks a population list from which to sample is to limit the goal of generalizability and construct a sample purposively. A sampling frame can be built by aggregating the entities known to, and judged to be relevant by, the investigators. Then, scholars can purposively sample from that set of known actors. While there are a variety of heuristics that one might employ for purposively selecting cases, many approaches require detailed knowledge of the area developed through experience or prior research.

In the social movements literature, della Porta and Mosca (2009) provide an exemplary article using this sampling design. They used their detailed knowledge of global justice movement (GJM) organizations to select a sample of thirty-five organizations operating in six European countries and transnationally that were "involved in the main initiatives of the GJM" (della Porta and Mosca 2009: 774), which they believe represented the overall contours of the GJM.[5] They argue that a random sample of GJM websites would be impossible because of the lack of a population list and therefore construct their sample purposively by focusing on cases that capture variation on explanatory variables (della Porta and Mosca 2009: 774). They defend this decision arguing: "since our case selection respected the principle that 'we must not search for those observations that fit (or do not fit) our *a priori* theory," we do feel confident that the selection choices did not bias the statistical correlations among the coded variables" (della Porta and Mosca 2009: 775; emphasis in original).

Although the authors are contrite in the methods and findings sections when they acknowledge that "ours is not a random sample and therefore cannot be considered representative of the composition of the GJM in each country" (della Porta and Mosca 2009: 774), their conclusions seem to strive for more general claims about how organizations and movements use ICTs. By the close of the article, they cast their findings in terms of larger likely trends and sets of relationships that they expect might hold in other settings (although they acknowledge that this would require testing in future research to confirm). This move toward more robust conclusions can only happen if a simplifying assumption holds true: the best sample that can be purposively produced will be close enough to the true population of GJM websites in the six countries to yield accurate population estimates.

This approach has a great deal of intuitive appeal. First, it is able to build on the detailed knowledge of movements in specific countries that experts hold. Second, moving away from needing a full population as a sampling frame entirely solves a significant practical dilemma. Third, if one's knowledge of a movement is extensive enough, it may well be that a purposive sample could, at times, approach a random sample in terms of approximating population features and relationships, especially if it were large enough. Finally, if one's primary concern is with relationships between variables, as opposed to descriptive frequencies, then purposive samples might still contain the same kinds of relationships found in the population, even if they don't provide very good point estimates for univariate population characteristics.

However, the risks of this approach are also clear. To the extent to which one is interested in univariate point estimates of population characteristics, or thinks that relationships may differ among segments of the population, purposive samples risk serious sampling error. Since the representativeness of point estimates and the generalizability of casual dynamics are often among the most important things for which researchers strive, this is a serious risk.

Della Porta and Mosca do report on point estimates for various characteristics, suggesting that estimating population characteristics is a goal of their work.

RANDOM ORGANIZATIONAL SAMPLES

Other scholars have been heavily interested in generalizability and thus have searched for alternative ways to identify populations. Stein's (2009) random organizational sampling is an exemplar in this regard. This method conceptualizes social movement organizations (SMOs) as the critical unit to observe and thus tries to identify a population list of SMOs. Put differently, it implicitly asserts that a sampling frame made of SMOs is an adequate substitute for a larger population list. Specifically, Stein uses the Encyclopedia of Associations (EA) as her sampling frame for six different movements. She then randomly samples from the organizations listed in EA. After identifying an SMO, she searched for its website and coded it. She argues this gives her a random sample of movement websites.

This method makes two important assumptions that allow her to resolve sampling design problems. First, it assumes that the actions of social movements online can be comprehensively understood through SMOs. This is not out of line with a large tradition that uses SMOs as a lens for understanding larger social movements and their dynamics (see Clemens and Minkoff 2004 for a review). But, as I argue below, this is a more problematic assumption to build into online research because a number of researchers have argued that a great deal of meaningful online social action is being organized outside of organizations (Benkler 2006; Earl and Kimport 2011; Shirky 2008). If one only seeks to understand how SMOs use the Web, not how movement actors of all sorts use the Web, this sampling frame is excellent. But, if scholars hope to speak about movements more broadly, this sampling frame is potentially quite limited. I treat the extent to which this may cause a problem as an empirical question when I compare and contrast the three approaches below.

This method's second assumption is about the adequacy of using the EA as a sampling frame even for SMOs. EA is certainly an excellent guide to large and well-known organizations and has been used in social movements research before (Minkoff 1995, 1997). However, the extent to which EA captures younger organizations or very small or subnational organizations (at least for the editions Stein appeared to have used) is unclear at best. I argue that researchers need to be careful about what populations they generalize to when using EA as a sampling frame—it is best at providing a sampling frame for large and well-established SMOs.

It is important to consider the theoretical implications that this more constrained sample may have. If EA represents a sampling frame biased toward large and well-established SMOs' websites, then we must ask whether large and well-established SMOs are likely to use the Web just as other actors would. If they are, then there is no problem with generalizing beyond the fold of large and well-established organizations. But, if there is reason to believe that they do not use technology in the same way, then scholars using this approach must be very circumspect about how far their conclusions can generalize.

Existing research suggests that there are reasons to believe that organizations established before the pervasive use of technology tend to use that technology differently than organizations that emerge once a technology is embedded in the environment (DiMaggio, Hargittai, Neuman, and Robinson 2001). This means that the set of organizations likely listed in EA is theoretically consequential—how these SMOs use the Web should be expected to differ from how other actors use the Web. This is of no consequence when researchers are careful about how far their findings can reach, but it is of great import when broader generalizations are desired.

HACKING USER BEHAVIOR: STUDYING "REACHABLE" WEBSITES

A third approach to the lack of a good population list that could serve as a sampling frame has been to try to harness assumptions about how users locate websites to manufacture a population list of "reachable" sites, which can then be sampled. Earl (Earl 2006; Earl, Kimport, Prieto, Rush, and Reynoso 2010) developed this approach using two simplifying assumptions. First, she argues that students of public protest do not need to identify all online content relevant to protest, but instead only need to identify content that could be located by a user who did not already know its location (i.e., the goal is to identify public online protest material). These public sites are referred to as "reachable sites." This implies that this method would be poorly suited for studying covert social movement activity, but so long as public activity is the primary interest, the assumption should be unproblematic. Second, Earl argues that based on prior research, one can make assumptions about how users tend to locate websites, which is largely through search (e.g., Google) or navigating links from sites they have already found. Fortunately, using Google mimics both of these behaviors: (1) it returns results that a user could gain from searching it (or many other search engines); and (2) GoogleBot, which finds Web-based material for Google to index and deliver through search, identifies new sites by crawling links from known sites. As the most extensive search engine available at the time of her study, using Google allowed Earl to identify websites run by large and small actors and by emerging and long-standing actors as well as popular and relatively unknown websites.[6]

Thus, if one is willing to accept a sampling frame made of reachable sites, instead of a list of all websites, then sampling from a frame that closely mirrors the population is possible. Earl does this by pretesting search terms related to the topics she is interested in and then deploying multiple search terms per topic (between 6-14 depending on the topic) in Google (see Earl 2006 and Earl and Kimport 2011 for this method to identify online instances of four protest tactics, and Earl et al. 2010 for this method to identify websites associated with twenty different social movements). She uses a SOAP-API to automatically query Google and receive 1000 results per search term.[7] She then concatenates the results from multiple searches, yielding 6,000-14,000 results. Because some searches return the same URLs, she discards duplicate URLs, reducing each population of reachable sites by varying amounts depending on the amount of overlap in search results. At that point, she argues that a population list of reachable sites has been identified, which can be used as a sampling frame for public websites and which can be randomly sampled.

Since random sampling ignores the ranking Google assigned the site (i.e., it doesn't matter if the site was ranked first or 1000th in the sampling procedure), the sample is not structured by Google's Page Rank algorithm to any discernible degree. Random sampling has a variety of advantages, including the ability to identify less popular websites, small websites, and sites that are not well connected; random sampling should select these sites at whatever incidence they occur in the population. Put differently, whatever high stakes gaming goes into making the first page of Google results will not affect randomly sampled results from large queries (although it would affect a study that only used top 10 or top 100 search results to populate a sampling frame or a study that purposively samples based on popularity).

There are a number of advantages to this overall sampling design. First, if one agrees with the two simplifying assumptions (i.e., the focus on reachability and a researcher's ability to recreate what is reachable at a given time), then the method offers a true random sample of reachable Web content. Second, this method makes no assumptions about the kind of actor that would publish movement-related material online. This means that Earl's samples would only focus on SMOs to the extent to which SMOs are making most of the relevant online material. But, this method would also allow one to identify activity happening outside of SMOs as well (and at whatever rate it is occurring in the empirical world). As the comparison of methods below reveals, this is actually a substantial advantage. Third, this method is robust

to changes in the primacy of different search engines because one could use the same methods in another search engine if a competitor outpaced Google.

METHODS

Verifying that any of the three sampling designs is robust is difficult in the abstract. Indeed, precisely the problem that created the need for alternative selection methods—the lack of a population list from which one could create a sampling frame—also means that one does not know the true population parameters. This means that one cannot *a priori* know if a new selection method drew a sample that closely resembles the population or not! But, one can treat the true population as unknown and still compare the different population estimates yielded by each method. In the rest of this article, I outline a method for comparing findings between the studies to determine how different the underlying sampling designs and findings are. Since I am able to compare data collected on roughly the same movements, over roughly the same time periods, it is possible to determine how differently these sampling designs operate in terms of the ultimate samples they produced.

Specifically, I evaluate these three sampling designs by adjusting Earl's data to approximate Stein's and della Porta and Mosca's data as closely as possible, and I then compare the findings. In order to understand this comparison, I first review major similarities and differences between the datasets and then discuss specific adjustments made to Earl's dataset to increase commensurability with the other sets.

General Similarities and Differences in the Datasets

A key advantage to choosing these three studies is that they include data on almost exactly the same movements while also exemplifying three competing sampling designs. Earl's project was the most expansive: she examined twenty different social movement areas, including five of the six movements that Stein studied (all but media reform) and the single movement that della Porta and Mosca studied, the GJM. Accordingly, for each comparison below, I report three figures from Earl's data: the twenty movement results, results from the five movements that are comparable with Stein's data, and results from the GJM, which are comparable with della Porta and Mosca's study.

The time coverage is almost exactly the same as well. Stein examined her cases in 2006, which is the same year from which Earl's data is derived. Unfortunately, della Porta and Mosca examined their sites in 2005 (della Porta and Mosca 2005) so some difference in findings might be attributable to that difference, not the sampling design. However, there is no reason to believe that a major sea change in online activism happened between 2005 and 2006, so any differences attributable to timing should be slight. Moreover, in the findings discussed below, della Porta and Mosca often reach similar findings to Stein, suggesting that timing failed to significantly affect the results.

There is more geographical diversity amongst the projects, but this is unlikely to be responsible for many of the key differences examined in this paper. Stein explicitly examined American organizations while della Porta and Mosca examined websites from organizations across six different European countries and organizations that were active transnationally. Earl's sampling design was not focused on a single country, but nonetheless captured predominately (although not exclusively) American websites. Since this makes data from Earl and Stein most comparable, if there are places where Earl's and Stein's datasets converge but differ substantially from della Porta and Mosca's, readers might reasonably suspect that country-level differences matter in those instances. However, where Stein and della Porta and Mosca's findings converge, despite differences in countries and in year, it is likely that these differences owe to more fundamental differences based on sampling design (i.e., their sampling frames are more similar to one another than to Earl's design).

It is important to note, though, one other difference that could affect the comparability of the sampling frames: Earl designed her reachable populations around a more holistic definition of the movement arena, including both sides of a movement struggle. Instead of trying to capture only pro-life websites, she designed search terms to capture both pro-life and pro-choice websites and she analyzed these sites together as "abortion politics" websites. In contrast, Stein and della Porta and Mosca only included pro-GJM groups in their sampling frames. This strategy would produce artifacts that need to be accounted for if movements and countermovements use the Web in fundamentally different ways, but research suggests that opposing movements are often structurally isomorphic (Meyer and Staggenborg 1996; Rohlinger 2006), so this is unlikely to influence findings.

Reconfiguring Earl's Data

In addition to producing findings from different subsets of movements, there are a few more adjustments that are needed to make Earl's dataset more comparable. Earl's sampling frame necessarily meant that she captured a much wider array of websites related to each movement area than the other sampling frames. For instance, a government website reviewing environmental regulations designed to protect air or water quality could be among the reachable sites in her population. Likewise, a "green" company that attempts to resolve environmental harms through market solutions could also be captured within her reachable population. Since these sites are not strictly advocacy sites, it is necessary to remove these types of sites to render her data more comparable. This is accomplished by removing any websites that did not advocate on the issue under which it was sampled. For instance, both sites above would be removed from the sample because they were not involved in environmental advocacy. Likewise, sites that provided news or information on environmental topics but did not take an advocacy position would be removed, as would websites that might have once advocated on a given topic (like the environment) but were not advocating a position on that particular issue at the time they were examined (e.g., the site was sampled from the environmental reachable population but was advocating a position on globalization, not the environment, when studied).

Coding Rules and Categories

All three projects collected their data using quantitative hand coding of content by a research team.[8] To compare coding protocols for specific variables, varying levels of detail are available for each project. For Stein's coding procedures, only those reported in published papers using these data are available. For della Porta and Mosca's data, there is a published paper and a publicly available, online codebook.[9] Both inform my understanding of their coding procedures. I have full access to Earl's coding procedures. Below, I explicitly outline Earl's operationalization of each analyzed variable and compare that to what appears in published accounts of the other studies and in the della Porta and Mosca codebook.

Earl coded a site as being SMO-sponsored when an organization with some offline indication of existence claimed to direct, maintain, or sponsor the website. Offline existence was required to rule out organizations in name only. For instance, there were sites that claimed to be run by organizations but upon further inspection were actually run by a single individual who adopted an organizational name to presumably increase the legitimacy of the site. Not having this requirement would have overestimated the true level of organizational involvement, although having this requirement does risk underestimating the true level of organizational involvement if small organizations have not established offline presences. However, qualitative experience from coding suggests that this did not result in significant undercoding of organizations. Stein operationally defined organizations as any group appearing in the EA. Della Porta and Mosca discuss their project as a study of SMOs in one report (della Porta and

Mosca 2005) but elsewhere (della Porta and Mosca 2009) frame their study as more broadly including "unions, leftist parties, and NGOs, as well as networks and grassroots groups" (p.775). I adopt the broader view and assume their study is of movement-relevant organizations, not just SMOs.

Two measures of offline protest can be drawn from Earl's data. First, a looser measure uses a dummy variable that measures the presence (through hosting or linking) of any content on offline protests, such as reports on past protests, invitations to attend upcoming events, or protest calendars that include offline events. Offline protest opportunities that triggered coding include: rallies, demonstrations, marches, vigils, pickets, civil disobedience, commemorative ceremonies with social movement significance, dramaturgical demonstrations (e.g., use of puppets in protest), lobbying, symbolic displays (e.g., cross burnings), collective violence, strikes, social movement press conferences, lawsuits, physical phone banks, offline fundraising events that are also meant to raise public consciousness (e.g., breast cancer walks but not everyday organizational fundraisers), or other clear offline forms of action not elsewhere classified (e.g., hunger strikes). A stricter measure of offline protest opportunities limits the dummy variable to marking the presence of actual upcoming or ongoing (e.g., repeating vigils) offline protest events in which site visitors are still "at risk" of participating. It is not clear what kinds of offline tactics triggered Stein's or della Porta and Mosca's coding of offline events.[10]

Two similar measures of online protest can also be created from Earl's data. The looser measure is a dummy variable capturing the presence (through hosting or linking) of online protest opportunities, such as online petitions or email campaigns. The stricter measure again restricts the dummy variable to noting the presence of upcoming or ongoing online protest in which the site visitor is still at risk of participating. Online protest opportunities that could trigger either dummy variable include: online petitions; send-a-message, emailing, twitter posting, or posting campaigns on social network sites; boycotts, letter-writing campaigns, phone call campaigns, and faxing campaigns organized online; various kinds of hacktivism; online fundraising campaigns designed to raise public consciousness (as was true for offline activism, this would not include everyday fundraisers); or other clear online forms of action not elsewhere classified (e.g., photo petitions).

Stein (2009: 759) provides some examples of online actions her team coded, including "online petitions, email campaigns, online action or surface letter-writing campaigns," but does not provide a comprehensive list of coded protest forms. Based on their published work and codebook, it appears that della Porta and Mosca (2009: 781) coded for online petitions, e-postcards, netstrikes (i.e., a distributed denial of service attack), and mailbombings (i.e., similar to a distributed denial of service attack but via overloading email systems). Neither project is explicit about whether actions must be upcoming or not, and thus it is not clear whether the looser or stricter measure from Earl's data is closer to Stein's or della Porta and Mosca's operationalization (although della Porta and Mosca's codebook does indicate they were coding a "call for" these activities). Therefore, findings using both measures are presented below.[11]

Stein and della Porta and Mosca published point estimates for several specific forms of online protest, including petitions, email-related campaigns, and various forms of hacktivism. For petitions, it is possible to provide comparable estimates across all three datasets. For messaging-related campaigns, it is less clear how comparable the measure of della Porta and Mosca is to the other measures. Della Porta and Mosca explicitly measured "e-postcards," which is not a terminology common in the U.S. Based on other work from the same project that provides an example of an e-postcard (della Porta and Mosca 2005), it seems likely that in substance an e-postcard is similar to an email or letter-writing campaign or a "send a message" button on a U.S. website. To build the most comparable measure possible, a dummy variable is coded as 1 if there is an email, send-a-message, or online letter-writing campaign using Earl's data. This matches several discrete variables that Stein coded, including email campaigns and surface letter-writing campaigns, but she reports each of those figures

separately. In terms of hacktivism, Stein reported on distributed denial of service (DDoS) attacks and della Porta and Mosca reported on netstrikes (which seem to be like DDoS attacks) and mailbombings (which seem to be like DDoS attacks on email addresses instead of on websites). Earl codes specifically for DDoS attacks but also codes for other forms of malicious code, hacktivism, and website defacement.

All three studies also measured the provision of information about the movement's issue on the website. To operationalize this from Earl's data, a dummy variable indicates whether or not the site provides (by hosting or linking) any reasonably verifiable information on the movement's focus. Information may have been represented through press accounts of an issue, research reports, agency or government agency reports, reports from nonprofits, or other forms of information that was not simply public relations but contained actual information. This variable seems comparable to della Porta and Mosca's (2009: 776) variable, which captured "political education via articles, papers or dossiers." The closest variable from Stein's study coded "project descriptions or news" (Stein 2009: 759).

Summary of Methodological Similarities and Differences between Studies

Overall, aside from the major differences in sampling designs, the differences between Earl's and Stein's studies are slight. While Earl does not capture data on the media reform movement, this should not be a major concern unless the media reform movement was an outlier that substantially influenced the overall rates that Stein reported across her six movements.

Design differences between these two studies and della Porta and Mosca's studies were larger: della Porta and Mosca used 2005 data while the other studies used 2006, and della Porta and Mosca drew on data from six different European countries while Earl had a largely American sample and Stein's was entirely American. That said, Stein's and della Porta and Mosca's findings converged on the three key findings discussed below, but diverged from Earl. This suggests that the year and country differences mattered less since Stein's findings still converged with those of della Porta and Mosca. Therefore, I argue that most of the differences below owe to differences in sampling design.

FINDINGS: HOW MUCH DOES SAMPLING DESIGN AFFECT RESULTS?

A comparison of the three research projects reveals substantial differences in findings, which I argue are attributable to differences in the sampling design. The single most important difference between the three projects is in the role of non-SMOs and/or non-organizations in online protest. Della Porta and Mosca structure their sampling frame around protest-relevant organizations (close to what many would think of as SMOs, but slightly more inclusive). Stein structured her sampling frame around SMOs. This means that neither study examined social movement websites created by non-organizations. Earl's sampling frame, though, allows for non-SMOs and even non-organizations to be content creators. As shown in table 1, she finds extensive involvement of non-SMOs: in 2006, she found that only 48 percent of the websites associated with all twenty social movements she studied were created by SMOs. But, that increases to 68 percent if you limit Earl's sample to websites from the five movements that overlap with Stein. The percentage skyrockets to 81 percent if you limit Earl's data to only GJM sites, which is the movement that overlaps between all three studies.

This finding suggests that, in addition to sampling design differences, movement selection is likely to impact the representativeness of findings for protest generally, or at least the generalizability to movements outside of those studied. It would appear, for instance, that GJM has a very different authorship pattern for websites than other movements, which is apparent in the contrast between Earl's twenty-movement percentage, the five-movement percentage, and the GJM percentage. This has an upside for della Porta and Mosca in that they

Table 1. Summary of Findings across Three Studies on Key Issues

	Stein	Della Porta and Mosca	Earl
SMO-run website	100%	100%	48% overall 68% with Stein sample 81% with GJM only
Offline protest through website (loose measure)	60%	67%	19% overall 19% with Stein sample 27% with GJM only
Offline protest through website (stricter measure)	60%	67%	8% overall 3% with Stein sample 8% with GJM only
Online protest through website (loose measure)	20%	Unclear but at least 33%	43% overall 56% with Stein sample 54% with GJM only
Online protest through website (loose measure)	20%	Unclear but at least 33%	37% overall 47% with Stein sample 50% with GJM only
Offline versus Online	Offline > Online	Offline > Online	Offline < Online
Hacktivism	Rare (1/89 cases)	Negligible (less than 15%)	No occurrences
Informing	More than 80%	90%	90% overall 99% with Stein sample 100% with GJM only

Note: Della Porta and Mosca focused on protest-relevant organizations, which is a slightly broader category compared to SMOs, and includes unions, leftist parties, NGOs, and grassroots groups, and networks.

they chose a movement for which a sampling frame that misses non-organizational websites affects them the least—they are only missing twenty percent of the online social movement activity that Earl's method was able to detect based on their sampling frame. But, it also suggests that their findings are unlikely to be representative of online dynamics for other movements since other movements are far more likely to have a lower percentage of SMO-operated sites according to Earl's findings.

Stein's sampling frame's reliance on SMOs for selection hurts her representativeness for five of the movements she studied: almost one-third of the online social movement activity that Earl was able to identify related to these five movements is excluded from observation by her design. But, the five movements she studied are closer to other movements that Earl studied.

Another big difference in findings between studies has to do with the kinds of protest activities supported through coded websites. As table 1 shows, Stein (2009: 759) found that about 60 percent of sites coordinated offline actions. Della Porta and Mosca (2009: 780) find that 60 percent of websites had calendars of upcoming offline events, 36 percent had other "information on offline forms of action," and "two-thirds of [the] web sites advertise the participation of their organization in [an offline] protest campaign." While it is unclear which of these measures is most comparable to Earl's measures of offline protest, the differences in findings are stark enough that coarse comparisons are possible. Earl finds that across all

twenty of the movement areas she studied, 19 percent of sites featured offline protest using her looser measure, compared with only 8 percent of sites using her stricter measure that requires events to still be open to new participants (see rows 2-3 in table 1 for the looser and stricter measures).[12] The findings are similar when the comparison is limited to the five movements that overlap with Stein's data: using the looser measure, still only 19 percent of websites featured offline protest, but the stricter measure is much lower at 3 percent of sites. The comparison is a little better when one limits Earl's sample to GJM websites: Earl's coarser measure suggests 27 percent of sites feature offline protest while 8 percent of sites offer visitors the option of learning about upcoming or ongoing offline protest. Looking across these values, it is clear that Earl's results are markedly different from the other studies in that she finds a bare minority of sites facilitating offline activism whereas the other authors find a majority of sites engaged in this activity. This difference fits with Earl et al.'s (2010) claims that research was systematically overestimating the prevalence of sites facilitating offline protest.

Stein and della Porta and Mosca also both find lower levels of online action than Earl's data reveals, as shown in table 1. Stein (2009: 759) found that less than 20 percent of sites offered online protest options. Della Porta and Mosca (2009: 781) find that

> online forms of action are promoted less often than offline tactics: almost 30 percent of the analysed web sites use the online petition; almost 18 percent propose to their users a form of online mobilization like the e-postcard; and 15 percent publish concrete information about online forms of action on the web site. The percentage is even lower if we consider the presence of calls to netstrikes and/or mailbombings; other forms of online mobilizations are much more widespread, although still limited to a minority of web sites.

From their reporting, it is difficult to determine the overall percentage of websites that offered any online action since, presumably, some of the sites that offered online petitions also offered other online mobilizations, but the highest single rate reported was 30 percent.

Earl finds higher levels of online protest actions (see table 1, rows 4-5 for looser and stricter measures). In her twenty-movement sample, 43 percent of websites offered online actions, using the looser definition, and 37 percent using the stricter conceptualization of ongoing or upcoming actions. But, when Earl's data is reduced to cover only the five movements that overlap with Stein, the figures rise to 56 percent of sites using the looser definition and 47 percent of sites using the stricter definition. The percentages stay high if the sample is limited to GJM websites: 54 percent of sites featured online protest opportunities using the looser definition and 50 percent had upcoming or ongoing online protest opportunities. As row six in table 1 demonstrates, Earl not only identified more online protest than other studies, but she also found that online protest was more common than offline protest, which is also in contrast to the other two studies.

In terms of specific tactics, online petitions made up a comparable or smaller share of all online protest in Earl's data than in della Porta and Mosca's data (depending on the comparison), but both studies found more online petitions than Stein's data identified. Stein (2009: 760) found that 11 percent of sites offered online petitions, while della Porta and Mosca (2009) found that roughly a quarter of sites featured petitions. Across all twenty movements Earl examined, about 16 percent of sites offered ongoing or upcoming online petitions. Twenty-four percent of the websites from the five movements that overlapped with Stein offered ongoing or upcoming online petitions in Earl's data. For GJM websites, Earl's data show about 23 percent of sites had ongoing or upcoming petitions.

Stein (2009: 760) found that 12 percent of sites offered email campaigns and 16 percent offered letter-writing campaigns. Della Porta and Mosca (2009: 781) found a comparable rate: they found that 18 percent of sites offered e-postcard campaigns. Earl's data show higher rates. For all twenty movements, Earl found that 28 percent of sites offered some kind of email or messaging-related campaign. For the five movements that overlapped with Stein's

study, Earl found that 42 percent of sites offered email or messaging-related campaigns. When looking only at GJM websites in Earl's data, about half of GJM websites had email or messaging-related campaigns.

It is important to stress the importance of these major differences in findings. The first difference—the rate of non-SMO driven online protest material—reveals a major empirical blind spot for any SMO-based, or even organizationally based, sampling frame. It suggests that anyone sampling online material based on an organizational frame needs to be very clear in their findings, discussion, and conclusion that the findings can only speak to what organizations do online, not to what a movement in general looks like online.

The second two differences—differences in the proportion of sites engaging in offline and online protest—are also critical. As Earl et al. (2010) showed, research on activism that is happening offline but being facilitated online has strikingly different theoretical expectations about the role of ICTs in protest than research examining activism that is occurring online. In the former, current theories tend to be supported with minor adjustments, but in the latter, pre-Internet theories usually struggle to explain online participation and organizing (Earl and Kimport 2011). Both Stein's and della Porta and Mosca's studies found that the majority of Internet activism was really offline action that was facilitated online, while very little genuinely online action was occurring. Earl et al. (2010) reported exactly the opposite. It would appear that these differences in large part owe to differences in the sampling design, most likely attributable to whether the sampling frame is limited to organizations.

In addition to these major differences, there were also more minor differences in findings. Stein (2009: 759), for instance, found that one of the 86 sites she studied engaged in a DDoS action (which is roughly a little more than 1 percent of her sample; see table 1). Della Porta and Mosca (2005) found 7 percent of sites engaged in mailbombing and 6 percent in netstrikes. Earl found no instance of this kind of action on any website associated with any of the twenty movements she studied. This is the only comparison on which Earl's data showed a lower incidence of a specific form of online activism.

It is hard to determine why the results vary so significantly on this point. One initial thought would be that because hacktivism is illegal in many countries, it is more covert. This would imply that Earl's method is less apt to capture sites discussing hacktivism because of her focus on public protest through reachable sites. But, on closer inspection, this rationale for the difference does not hold: unless the entire organization was clandestine, Earl's method is as well equipped to identify SMO's websites as the other designs. Since Stein's and Earl's findings are closer on this estimate, it is also possible that this discrepancy has to do with geography: it is possible that European organizations are more likely to engage in hacktivism. Alternatively, della Porta and Mosca's purposive sample included anarchist organizations, which have been associated with these tactics. Thus, it may be that by including more radical organizations in their sample, they increased the likelihood of finding hacktivist tactics.

There were also places where the samples produced overlapping findings. Stein (2009) reported that an overwhelming percentage of websites included news or information on their causes (see table 1). Similarly, della Porta and Mosca (2009: 776) report that 90 percent of the websites they studied offered information on their causes. Earl found that 90 percent of the websites associated with all twenty movements she studied, 99 percent of the websites associated with the five movements that overlapped with Stein's research, and 100 percent of GJM websites provided information related to their cause. Given that the specific indicators that each study used do not perfectly align for this measure, it is impossible to give a strict comparison across these values. But, all reach the same general conclusion: websites are exceedingly likely to offer information on the causes they attempt to forward. This suggests that for variables that show very little variation across websites because of exceedingly high adoption rates, sampling design will have a minimal impact on findings.

DISCUSSION AND CONCLUSION

I began this article by noting that a major issue facing students of online activism is the lack of a population map of major features of online protest—it is currently very difficult to situate individual case studies, or studies of even a few movements, into a larger social space. Methodologically, this means it is very hard to know how representative any given case is or how representative any given set of cases is. This has real implications for the state of our collective knowledge. For instance, despite extensive research on the Zapatistas' use of the Internet, it is not at all clear whether the ways that the Zapatistas used the Internet were empirically rare or common. It is impossible to know whether divergent findings from two or more studies on the same movement owe to differences in how cases were selected or to other factors, and to know which of the divergent studies tracks more closely to population characteristics.

The traditional solution to such a problem is simply to build a map of population characteristics by surveying a random sample of the population. This is how we know what the average Fortune 500 company pays its CEO, how many Americans have participated in protests during their lifetime, and how frequently national SMOs form coalitions with one another. But, for students of online protest, the lack of a suitable population list from which to draw a sample makes this standard route intractable at first glance.

Scholars have therefore struggled to identify the best alternative routes to selecting cases to study. One direction has been to harness detailed, substantive knowledge of a movement—to really engage the expertise of the researcher—and select cases purposively. Della Porta and Mosca (2009) offer an excellent example of this approach with their study of websites associated with protest-relevant organizations in the GJM. They argue that because random samples of online materials are impossible to produce, expert knowledge in constructing a sampling frame and in executing the sampling itself is the next best route. In doing so, they implicitly hope that the set of sites they select are not so different from key population characteristics as to make their data idiosyncratic.

Another approach, represented by Stein (2009), uses a sampling frame focused on SMOs to identify candidate movement websites. The sampling frame makes methodological dilemmas much easier to overcome because (better and worse) population lists of organizations exist. Specifically, Stein used the Encyclopedia of Associations as her sampling frame, which has been used by protest scholars before and is likely to capture established, national SMOs. Solving this methodological dilemma, though, comes at a price: this method ensures that relevant protest activity that is happening outside of SMOs will not be observed. If SMOs make up the entirety of a protest sector, this will not matter, but if a non-negligible amount of online protest activity is organized outside of SMOs, this will be a more substantial issue (unless, of course, one refuses to generalize beyond SMO activity). However, in terms of sampling method, the use of random sampling increases the odds that the sample will at least be representative of SMOs operating across the movements she studied.

Earl offers a different approach to sampling. She argues that while it is impossible, even for companies as large and well resourced as Google or Microsoft, to catalog all Web-based material, it is possible to identify the set of sites that an average user could be at risk for finding online without having a direct URL. She refers to sites that users can find through search or through links from known sites as "reachable sites" and argues that search engines such as Google can be used for repetitive and overlapping searches that, when concatenated, produce a population of reachable sites that can be used as a comprehensive sampling frame. While this assumption does not make sense for scholars interested in covert activity, it does offer a strong sampling frame for scholars interested in public protest online.

The objective of this paper is to lay out these competing approaches and their assumptions, and then use comparisons between the findings from exemplar studies using each design to demonstrate and explore the effects of sampling design. Because the three cases in my

analysis represent broad approaches to solving the sampling dilemma in online activism research, I am able to draw more general conclusions about the effects of sampling. There are five key findings. First, and most importantly, studies that rely on organizations to build their sampling frame (and therefore neglect websites produced by a wider variety of actors) miss a great deal of online action, although the extent to which this is a problem varies by movement. Second, organizationally focused sampling frames, whether sampled randomly or purposively, tend to underestimate the amount of online forms of activism when compared to samples drawn from sampling frames that include both SMO-produced and non-SMO sites. Third, and by contrast, organizationally focused sampling frames, whether sampled randomly or purposively, tend to overestimate the amount of online facilitation of offline protest when compared to samples from sampling frames that include both SMO-produced and non-SMO-produced sites. Fourth, there are differences in the estimated frequencies of specific kinds of actions, including the frequency of online petitioning, online messaging campaigns (emails, letter-writing, e-postcards), and hacktivism. While differences in the estimated frequency of petitioning and messaging likely depend on whether the sampling frame is limited to organizationally produced sites or not, differences in findings related to the frequency of hacktivism seem to largely turn on whether random and purposive sampling methods are used. Lastly, sampling design decisions do not seem to influence estimates of very pervasive types of activities such as sharing news and information about a cause.

To cast these findings in a different light, on most points (with the exception of the finding on hacktivism), the studies that share similar sampling frames have similar findings, even though the sampling method varies between them: both studies inherently ignore action not organized by organizations, find high levels of online support for offline protests, and find much lower levels of actual online protest (e.g., emailing campaigns). This suggests that if one is interested in what organizations are doing online, although random sampling is generally preferable from a hypothetical standpoint, it might not be as empirically consequential whether one randomly samples or harnesses expert knowledge to construct a wide and varied sample.

In terms of translating these findings into best practices, I suggest a first principal question researchers need to ask themselves and orient their discussion of findings and implications around: *Is this a study of SMOs (or protest-relevant organizations) and their activities or a study of a broader array of protest-relevant actors?* This is both a theoretical and practical issue. At a theoretical level, this question requires scholars to consider and defend how they conceptualize social movements more broadly and to assess how well their operationalization represents that conceptualization. For instance, a scholar may claim that organizations are the most meaningful aspect of social movements, and thus the only type of agent that merits study. It is certainly the case that a large number of scholars have conceptualized and operationalized movements in this way offline. But, this should require specific theoretical attention and care for the decision's methodological consequences. Alternatively, one might conceptualize a movement as the aggregate of actions taken towards its goals, even if by disconnected or relatively unorganized individuals. An offline analog to this conceptualization uses protest events to study movements. Again, this requires specific attention and consideration.

At a practical level, if a sampling frame is SMO-focused, or organizationally focused, findings presented here suggest that researchers with sufficient movement expertise may be able to purposively sample from a sampling frame based on groups with which they are familiar. However, researchers with less movement expertise should still search for, and use, the best available population lists as sampling frames (e.g., the Encyclopedia of Associations).

In contrast to the organizationally focused studies, the sampling frame focused on reachable websites finds significant levels of social movement activity happening outside of SMOs, lower levels of online support for offline protest, and higher levels of online protest. These are all major and very consequential differences in findings when compared to the other two studies. This, in turn, suggests that if researchers are interested in generalizing to a broader

array of protest-relevant actors than SMOs, or want to speak about movements more generally, then they should prefer a method like Earl's reachable population method.

Of course, there are two exceptions to these general guidelines that scholars need to consider. First, when scholars are studying actions that have very low incidence rates, random sampling methods are likely to do a poor job in identifying these rare cases. In principal, a random sample should only capture cases at roughly the frequency in which they occur in the population. So, if something occurs only one out of every 10,000 instances in the population, a researcher might or might not observe it in a sample of 100, 500, or 1000 cases, for instance. Stein did observe one instance of hacktivism in her sample of 89 organization's Web presence, but Earl did not observe a single instance across thousands of observations. Della Porta and Mosca purposively sampled and in doing so might have been able to include deviant cases more easily. They found a higher incidence of hacktivism. I would argue that when individuals are interested in rare cases, they should consider case-study-based approaches so that they can maximize their knowledge of rare cases and spend their research energies on understanding these cases instead of tracking the rarity of such cases.

Second, the comparisons demonstrate something else that sampling theory suggests might theoretically be true. When a characteristic is saturated across the population such that virtually every population member has that characteristic, virtually any draw from the population, however it is structured, will return an adequate estimate. For instance, because nearly all websites provide information on the issues related to their protest activity, the three sampling methods produced very similar findings for this measure. The rate of this practice was so high that each of the three sampling designs can detect it and detect it at approximately the same rate. But, we should only expect this to be true for endemic characteristics.

In sum, the comparison of these approaches suggests that a major switching point in methodological decision making is whether or not the sampling frame includes nonorganizational content creators and, hence, whether SMO and/or organizational action versus social movement activity more broadly is the focus of study. Researchers also need to consider the expected prevalence rate of characteristics they intend to study, because high incidence characteristics appear relatively unaffected by sampling design and very low incidence characteristics are poorly studied via probabilistic sampling methods. Of course, other resolutions to the methodological dilemma discussed here might be invented, and it will be up to future research to evaluate that method against other existing solutions.

NOTES

[1] I am using sampling design to cover all elements involved in sampling. The sampling frame is the pool of cases that investigators hope represents the population closely and is the pool from which cases will be drawn. Thus, its construction is of considerable theoretical importance because how well or poorly it approximates a full population list is critical to determining representativeness. The sampling method is how cases are selected from the sampling frame (e.g., random, purposive).

[2] I acknowledge that one major issue involved in research on both online and offline activism is how exactly one conceptualizes and operationalizes a movement. In offline research, for instance, scholars have argued about whether tracking social movement organizations, protest events, or other metrics are best at capturing the trajectories of movements. Likewise, online scholars must decide how they bound what constitutes part of a movement. This paper takes an empirical look at the consequence of bounding movement activity online in two different ways: via organizations (as in the first two studies discussed) versus via activity (as in Earl's study).

[3] While IP addresses, which are critical to the underlying infrastructure of the Web, are patterned, they correspond to higher order domains. A great deal of material may be associated with a domain, including material by distinct authors. Moreover, IP addresses are not clustered by content type.

[4] While computational advances may allow for efficient processing of such large-scale data in the future, it remains a serious research dilemma today.

[5] Della Porta and Mosca argue that the GJM includes organizations dedicated solely to the GJM as well as organizations that take up issues related to the GJM, such as unions and human rights activists (see della Porta and Mosca 2009: 774).

[6] Earl's sampling frame should include resource-poor websites, websites produced by emerging actors, and websites that are not very well known at whatever rate they occur in the overall population. By contrast, it is unlikely that a

sampling frame based on the EA will include resource-poor groups, emerging groups, or relatively obscure groups. Likewise, since della Porta and Mosca include only organizations known to them and their research team in their sampling frame, they are also less likely to include resource-poor groups, emerging groups, or relatively obscure groups in their sampling frame.

[7] Google has subsequently discontinued the SOAP-API that Earl used, but it is still possible to do these searches manually. However, one must take steps to anonymize their machine so that Google cannot tailor results, which would make the searches unreplicable by other scholars.

[8] The reported reliability for the manual coding did vary significantly by project: della Porta and Mosca (2009: 775) reported reliability scores above 50 percent for each variable, and Stein (2009: 758) reported Krippendorf alpha scores of higher than .6 (ranging up to 1.0). Earl reported overall scores above 90 percent. However, it is unlikely that differing levels of reliability are responsible for the differences in findings examined in the paper.

[9] See http://demos.eui.eu/PDFfiles/Instruments/wp2codebook_final.pdf.

[10] Della Porta and Mosca's codebook does query whether the website has an "action/event calendar," but the codebook does not define what an action or event would include, which makes it difficult to make an entirely comparable measure using Earl's data.

[11] It is important to note that online protest opportunities might theoretically be more likely to be coded as ongoing or upcoming because Earl coded this variable based on how it was represented on the site. This means that if the site was still set up to facilitate a campaign and did not say the campaign was only active until a certain date, Earl coded that action as ongoing. This would be true even if a politically knowledgeable individual would recognize the issue was stale. Because offline protests are almost always associated with a specific date, "stale" events were more often coded as past events since they were not represented as still being open for participation. But, this is of little actual consequence in the analysis since findings using the coarser measure for offline and online participation still differ dramatically between studies, and differ in the same way that the stricter measures do.

[12] This percentage does not map exactly onto figures from Earl et al. (2010), because the number reported above doesn't differentiate between sites that *only* offered offline protest opportunities versus sites that offered *both* offline and online protest opportunities.

REFERENCES

Ackland, Robert. 2009. "Social Network Services as Data Sources and Platforms for e-Researching Social Networks." *Social Science Computer Review* 27(4): 481-92.

Ayres, Jeffrey M. 1999. "From the Streets to the Internet: The Cyber-Diffusion of Contention." *The Annals of the American Academy of Political and Social Science* 566(1): 132-143.

Benkler, Yochai. 2006. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. New Haven, CT: Yale University Press.

Bennett, Daniel, and Pam Fielding. 1999. *The Net Effect: How Cyberadvocacy is Changing the Political Landscape*. Merrifield, VA: e-advocates Press.

Carty, Victoria. 2002. "Technology and Counter-hegemonic Movements: The Case of Nike Corporation." *Social Movement Studies* 1(2): 129-146.

Clemens, Elisabeth S., and Debra C. Minkoff. 2004. "Beyond the Iron Law: Rethinking the Place of Organizations in Social Movement Research." Pp. 155-170 in *The Wiley-Blackwell Companion to Social Movements*, edited by David A. Snow, Sarah A. Soule, and Hanspeter Kriesi. Malden, MA: Blackwell Publishing.

della Porta, Donatella, and Lorenzo Mosca. 2005. "'Searching the Net' Report Prepared for the Wp2 of the Demos Project—Democracy in Europe and the Mobilization of the Society." Firenze, Italy: European University Institute.

———. 2009. "Searching the Net: Web Sites' Qualities in the Global Justice Movement." *Information, Communication & Society* 12(6): 771-92.

DiMaggio, Paul, Eszter Hargittai, W. Russell Neuman, and John P. Robinson. 2001. "The Social Implications of the Internet." *Annual Review of Sociology* 27: 307-36.

Earl, Jennifer. 2006. "Pursuing Social Change Online: The Use of Four Protest Tactics on the Internet." *Social Science Computer Review* 24(3): 362-77.

Earl, Jennifer, and Katrina Kimport. 2011. *Digitally Enabled Social Change: Activism in the Internet Age*. Cambridge, MA: MIT Press.

Earl, Jennifer, Katrina Kimport, Greg Prieto, Carly Rush, and Kimberly Reynoso. 2010. "Changing the World One Webpage at a Time: Conceptualizing and Explaining 'Internet Activism.'" *Mobilization* 15(4): 425-46.

Earl, Jennifer, and Alan Schussman. 2003. "The New Site of Activism: On-Line Organizations, Movement Entrepreneurs, and the Changing Location of Social Movement Decision-Making." *Research in Social Movements, Conflicts, and Change* 24: 155-187.

———. 2004. "Cease and Desist: Repression, Strategic Voting and the 2000 Presidential Election." *Mobilization* 9(2): 181-202.

Garrido, Maria, and Alexander Halavais. 2003. "Mapping Networks of Support for the Zapatista Movement: Applying Social-Networks Analysis to Study Contemporary Social Movements." Pp. 165-184 in *Cyberactivism: Online Activism in Theory and Practice*, edited by Martha McCaughey and Michael D. Ayers. New York: Routledge.

Gurak, Laura J. 1999. "The Promise and the Peril of Social Action in Cyberspace." Pp. 243-63 in *Communities in Cyberspace*, edited by Marc A. Smith and Peter Kollock. London: Routledge.

———. 1997. *Persuasion and Privacy in Cyberspace: The Online Protests over Lotus MarketPlace and the Clipper Chip*. New Haven, CT: Yale University Press.

Gurak, Laura J., and John Logie. 2003. "Internet Protests, from Text to Web." Pp. 25-46 in *Cyberactivism: Online Activism and Theory and Practice*, edited by Martha McCaughey and Michael D. Ayers. New York: Routledge.

Martinez-Torres, Maria Elena. 2001. "Civil Society, the Internet, and the Zapatistas." *Peace Review* 13(3): 347-55.

Meyer, David, and Suzanne Staggenborg. 1996. "Movements, Countermovements, and the Structure of Political Opportunity." *American Journal of Sociology* 101(6): 1628-60.

Minkoff, Debra C. 1995. "Interorganizational Influences on the Founding of African American Organizations, 1955-1985." *Sociological Forum* 10(1): 51-79.

———. 1997. "The Sequencing of Social Movements." *American Sociological Review* 62(5): 779-99.

Rohlinger, Deana A. 2006. "Friends and Foes: Media, Politics, and Tactics in the Abortion War." *Social Problems* 53(4): 537-61.

Schussman, Alan and Jennifer Earl. 2004. "From Barricades to Firewalls? Strategic Voting and Social Movement Leadership in the Internet Age." *Sociological Inquiry* 74(4): 439-63.

Shirky, Clay. 2008. *Here Comes Everybody: The Power of Organizing Without Organizations*. New York: Penguin Press.

Smith, Peter J., and Elizabeth Smyth. 2001. "Globalization, Citizenship and Technology: The Multilateral Agreement on Investment Meets the Internet." Pp. 183-206 in *Culture and Politics in the Information Age: A New Politics?*, edited by Frank Webster. London: Routledge.

Stein, Laura. 2009. "Social Movement Web Use in Theory and Practice: A Content Analysis of US Movement Websites." *New Media and Society* 11(5): 749-71.

Thelwall, Michael Arijan. 2004. *Link Analysis: An Information Science Approach*. San Diego, CA: Emerald Publishing Group.

Van Aelst, Peter, and Stefaan Walgrave. 2004. "New Media, New Movements? The Role of the Internet in Shaping the 'Anti-Globalization' Movement." Pp. 87-108 in *Cyberprotest: New Media, Citizens and Social Movements*, edited by Wim van de Donk, Brian D. Loader, Paul G. Nixon, and Dieter. Rucht. New York: Routledge.

Wray, Stefan. 1999. "On Electronic Civil Disobedience." *Peace Review* 11(1): 107-112.