Research Article

Sociological Methods & Research
1-56

DOI: 10.1177/0049124120914926
journals.sagepub.com/home/smr
SAGE

Using Location Data From Mobile Phones to Study Participation in Mass Protests

Assaf Rotman[1] and Michael Shalev[2]

Abstract

Automatically collected behavioral data on the location of users of mobile phones offer an unprecedented opportunity to measure mobilization in mass protests, while simultaneously expanding the range of researchable questions. Location data not only improve estimation of the number and composition of participants in large demonstrations. Thanks to high spatial and temporal resolution they also reveal when, where, and with whom different sociopolitical sectors join a protest campaign. This article compares the features and advantages of this type of data with other methods of measuring who participates in street protests. The steps in preparing a usable data set are explained with reference to a six-week campaign of mass mobilization in Israel in 2011. Findings based on the Israeli data set illustrate a wide range of potential applications, pertaining to both the determinants and consequences of protest participation. Limitations of mobile location data and the privacy issues it raises are also discussed.

Keywords

social movements, mobile location data, mass protests, Israel, big data

[1] Tel Aviv University, Israel
[2] The Hebrew University of Jerusalem, Israel

Corresponding Author:
Michael Shalev, The Hebrew University of Jerusalem, Mount Scopus, Jerusalem 91905, Israel.
Email: michael.shalev@gmail.com

On January 20, 2017, masses of well-wishers gathered at the National Mall in Washington DC to be present at Donald Trump's inauguration. The following day an even larger crowd appeared at the same venue for the Women's March, a protest against the new president. The *New York Times* called on the services of two "crowd scientists" from Britain to compare how big these crowds actually were. After studying still and video images of both events, the experts created a map divided into segments with different crowd densities. They calculated that at least 470,000 people were present at the peak of the Women's March, approximately three times their estimate for the Trump inauguration using the same method (Wallace and Parlapiano 2017).

Later that year Fysical, a commercial vendor of data culled from mobile phones, reported the same 3:1 ratio (Mann 2017). However, instead of manually evaluating crowd density based on images generated when attendance reached its peak, the company relied on stored data collected by applications that openly or surreptitiously monitor the movement of smartphones in real time. A similar promotional report from Safegraph, a rival vendor of location data, went further than this (Fox and Gu 2017). By merging information on sightings of mobile phones with Census Bureau data about the localities where the owners of these devices appeared to reside, data analysts were able to estimate not only how many people were present at each event but also what kinds of localities they came from. They found, for example, that the people in the Inauguration crowd were more likely to come from the South, and less likely to live in affluent communities, than their counterparts at the Women's March.

The growing availability of continuously recorded data on the location of mobile phones opens up new vistas for studying large-scale street protests. In the field of social movement research, when addressing existing questions, this type of big data can complement and in some respects even substitute for existing methods. More than this, it becomes possible to address questions that were rarely or never asked before, simply because suitable data were not available. Scholars interested in taking advantage of the new sources need to know the kinds of research questions that location data can address, the types and quality of available data, and what is involved in their processing and analysis. This article contributes to meeting these needs, with the aim not only of facilitating but also inspiring future research.

To illustrate both methodological and analytical aspects of using location data from mobile phones to study large demonstrations, we focus on a protest campaign[1] in Israel in summer 2011. This six-week-long series of demonstrations, similar in scale and form to the better known *Indignados* campaigns that preceded it in Spain and Greece, is believed to have mobilized one fifth of Israel's adult population (Perugorría, Shalev, and Tejerina 2016). In order to analyze participation in the protest, we acquired a data set derived from cellular phone signals transmitted from the locations of protest events in real time. This data set records the number of devices sighted at each event over the course of the campaign and their distribution between virtually all residential census tracts in the country. Publicly available information on the attributes of census tracts (towns and urban neighborhoods) was then merged with the participation data. This type of data set provides researchers with a new basis for examining the social and political coalitions that underlie mass mobilizations and for probing their spatial and temporal dynamics.

Improving our knowledge of who participates in large-scale street protests is important because even though demonstrations attracting tens or hundreds of thousands of participants are comparatively rare, they account for the lion's share of total participation (Beyerlein et al. 2016; Biggs 2016:22; McCarthy, McPhail, and Smith 1996). A more specific motivation for studying who participates in mass demonstrations is the unexpected appearance beginning in 2010 of a wave of mega-sized national protest campaigns, a phenomenon which some scholars believe was enabled by a new "connective" logic of collective action that is here to stay (Bennett and Segerberg 2013). These campaigns include the Arab Spring uprisings, anti-austerity protests in Iceland and Southern Europe in the wake of the 2008 financial crisis, protests against "stolen elections" and unresponsive governments in Eastern Europe and the former Soviet Union, and other instances of sustained mass mobilizations in countries as far afield as Turkey, Israel, Argentina, and Brazil (Ancelovici, Dufour, and Nez 2016; Della Porta 2015, 2017; Flesher Fominaya 2014; Gerbaudo 2017). Common denominators include the exceptional *scale* of the most prominent demonstrations, with participants sometimes numbering in the hundreds of thousands; the *persistence* of protest activities over a period of weeks or even months; and their geographical *dispersion* beyond the focal site at which the kickoff event took place. Reflecting these features, scholars have plausibly claimed that recent mass protests were unusually likely to mobilize "previously disengaged 'ordinary' citizens" (Onuch 2014:3) and correspondingly successful in assembling unusually diverse coalitions of participants (Della Porta 2015; Gerbaudo 2017; Goldstone 2011).

The comprehensive behavioral data generated by routine tracking of the location of mobile phones are particularly valuable for addressing two key empirical questions posed by mass protest campaigns of this type. (1) How encompassing was mobilization on the ground? Which sectors of society were more likely to participate and which tended to stay away? (2) To the extent that a campaign succeeded in activating a diverse coalition, did different classes, identity groups and political camps protest together (at the same times and places) or apart? A newly researchable question emerges from these twin concerns: *Who mobilizes when, where, and with whom*? The answers are capable of revealing the social structures and relations embedded in a chain of street protests.

As stated, the goal of this article is not only to describe the new data and the practical steps involved in making it usable (part 1) but also to concretely illustrate the types of empirical analysis and insights that it makes possible (part 2). Part 3 concludes with a summary of the potential uses to which this type of data can be put in researching large protests and a discussion of methodological limitations and ethical dilemmas.

Part 1: Measuring Protest Mobilization

This part of the article highlights the advantages of location data in comparison with alternative ways of measuring and analyzing the magnitude, composition, and variability of participation in mass protests. In addition, it explains how cellular networks and smartphone apps collect location data, how our data set and measures were constructed, and the types of analysis they make possible. We conclude with a summary comparison between the new methodology introduced here and existing alternatives.

The most noteworthy feature of protest participation data derived from automatic tracking of the location of mobile phones is that the data represent *observed behavior* that is monitored in *real time*. This makes it possible to avoid limitations and biases inherent in traditional methods, which rely on *subjective self-reporting* and/or gather information *retrospectively* in a context that inevitably differs from the moment of participation. Moreover, because location data are usually stored by vendors for several years, researchers have the luxury of being able to investigate unanticipated mass mobilizations after the event.

In what follows, three additional advantages are also emphasized. First, compared to other methods dependent on sampling, the samples available through mobile phone networks are far larger and also more likely to include "hard-to-reach" populations. Second, while estimates of the size of events based on location data are naturally of interest, the particular strength of this type of data is the ability to shed light on the sociologically and politically significant question of who joins the crowd and with whom. Third, for the purpose of studying protest campaigns (as opposed to one-off events), the high temporal and geographical resolution of location data offers a unique window onto the time–space dynamics of protest participation. These strengths should be borne in mind as we review alternative ways of quantitatively studying participation in large-scale street protests.

Overview and Comparison of Measurement Methods

Quantitative data on protest participants are traditionally gathered in one of the three ways: (1) estimates by journalists, organizers, and the authorities (usually the police); (2) retrospective self-reports by a representative sample of the population in response to questions posed in surveys; or (3) on-site surveys based on sampling demonstrators in the course of a protest event. Each method has strengths and weaknesses depending on the type of protest and the aim of the research.

Protest event analysis (PEA). PEA, the most established approach to quantifying protest activity, has many uses despite acknowledged weaknesses including lack of standards and various forms of selection bias (e.g., Hutter 2014). For the purposes of counting and classifying protest participants, the sources on which protest event analysts typically rely are of limited value. Size estimates gathered from these sources are inaccurate because of objective difficulties, the biases of those who furnish the estimates, and pressure to summon up a precise number. They are useful if the goal is simply to distinguish between orders of magnitude (Biggs 2016). But when it comes to probing the social and political characteristics of protest participants, media reports are inherently selective and impressionistic. They are particularly problematic for "open" protests that appeal to diverse participants without mediation by distinct organizations and networks (Walgrave and Klandermans 2010).

Population surveys. Nationally representative population surveys that ask respondents whether they participated in a specific event or campaign become feasible when a large fraction of the citizenry is mobilized. Although their use in studying mass protests dates back to the May 1968 events in France (Converse and Pierce 1986), population surveys have come into their own since 2010 and have been used to study participation in the Arab Spring, anti-austerity protests in Europe, and other large-scale campaigns worldwide.[2] These studies have made major advances in profiling participants and have shed much needed light on the motivations, practices, and mechanisms of mobilization underlying contemporary mass protests. Surveys are also attractive because they do not necessitate an instantaneous research effort, can be planned and carried out ex post, and may be used to sample all those who participated over the course of a protest campaign. That said, population surveys also have potentially grave limitations that have not always been acknowledged.

First, surveys that are truly nationally representative are difficult to obtain. Younger and low-income people are typically undersampled as are minorities and other “hard-to-reach populations” (Marpsat and Razafindratsima 2010). Participation by different social groups is likely to be inflated for some and underestimated for others. These limitations could be grave for studies relying on population surveys to evaluate the inclusiveness of recent mass protests.

In addition to problems of coverage, the retrospective self-reports on which population surveys rely are subject to hindsight bias and social desirability bias (Huber and Power 1985; Karp and Brockington 2005). Moreover, because protests often have feedback effects on citizens’ partisan affiliations and policy positions, it has been shown that the answers to questions asked during and after participation are an unreliable guide to protesters’ motives ex ante (Pierce and Converse 1990).

In the context of protest campaigns comprised of spatially and temporally varying episodes of contention, a final limitation of population surveys is that respondents may not recall exactly when and where they took to the streets. Even if such questions could be answered, extremely large samples would be needed in order to reach the level of disaggregation required for studying the dynamics of large-scale campaigns.

On-site surveys. On-site surveys are typically based on short interviews with protesters in situ, distributing questionnaires to participants in the hope that they will be filled in and returned afterward or a combination of the two (Andretta and Della Porta 2014; Walgrave and Verhulst 2008). Systematic sampling of participants is a serious challenge, but standardized protocols have been developed that, when properly followed, generate reliable samples (Walgrave and Verhulst 2011). Not surprisingly, however, the stricter the design requirements, the more difficult they are to meet in practice (Walgrave, Wouters, and Ketelaars 2016). Furthermore, rigorous on-site surveys generally impose high requirements in terms of funding, training, and forward planning, and the number of respondents who can be canvassed or interviewed with this method is limited (an average of around 225 in the 51 studies reviewed by Walgrave and colleagues). Finally, the advantages of rich individual data on protesters collected in or close to real time are accompanied by a grave limitation—the lack of comparable data on nonprotesters.

New technology-based methods. In addition to mobile phone location data, our focus here, other technically advanced approaches could potentially be used to estimate the size and/or composition of large demonstrations. Automated crowd counting based on analyzing images is a rapidly developing field, but the challenges posed by crowds that are large, dense, and moving have yet to be resolved (Henke 2016). “Soft biometrics” has spawned a number of automated systems for identifying the age, gender, and/or ethnicity of faces captured on video (Aziz et al. 2018; Dantcheva, Elia, and Ross 2016) but is only beginning to become feasible for events like large demonstrations (Joo and Steinert-Threlkeld 2018).

In contrast, studies of social movements based on data generated by social media are rapidly proliferating (e.g., Mobilizing Ideas 2015; Steinert-Threlkeld 2018). This body of research sheds light on the dual role of online protest activity, as both a separate sphere of contentious political action and a means of communication, coordination, and content creation that spurs and supports traditional offline protest. However, for present purposes, the question is whether the big data generated by social media can reveal the time and place in which street protests take place and relevant attributes of their participants. Judged by this standard, it can be argued that Twitter data, the basis of most quantitative analyses of social media and protest, have significant limitations.[3]

Identifying Crowds via Mobile Phones: Introducing the Data

Market researchers, transportation planners, and other applied researchers, as well as computer scientists and a sprinkling of academic social scientists, are increasingly utilizing data derived from the location of mobile phones (Calabrese, Ferrari, and Blondel 2014). This article uses the term “mobile location data” to refer to information on the location of mobile devices that is generated and stored by cellular networks and an increasing number of smartphone apps. Phone owners generally play a passive role and are typically unaware of the extent to which their location is being tracked, reused, resold, and integrated with other information about them.

Sources of mobile location data. There are two broad “families” of location data. The first and most established is data generated automatically by the internal operating procedures of cellular networks (Ahas 2011; Smoreda, Olteanu-Raimond, and Couronné 2013; von Mörner 2017).

Whenever mobile phones are turned on and connected to a cellular network, they emit signals that make them visible to nearby cell towers. As a switched-on device moves around, its sightings are monitored in order for the network to correctly route phone or data services to or from the device when needed. As a by-product, the approximate location of all devices that are connected to the network is continuously updated and recorded. This *passive monitoring* is what generated the data described and analyzed in this article. Another standard procedure, the call detail record (CDR)—designed for billing purposes—is *triggered by device activity*. It records information about each call, text, and data transfer sent or received by a mobile device, including the location of the tower with which the device communicated. There are numerous technical and substantive differences between location data derived from passive monitoring and device activity. Which is superior depends on the aims of the researcher and what type of data is available.[4]

Alternatively—and increasingly—location data are collected by mobile applications with access to the "location services" (in full, location-based services) built into smartphones (Mobile Marketing Association 2015; Valentino-Devries et al. 2018; Zickuhr 2013). Within this second family, information on the geographical positioning of a smartphone is obtained directly via GPS and indirectly by querying the location of nearby cell towers and Wi-Fi access points. Location services are routinely invoked by navigation apps or when users of social media or shopping apps choose to "check in." In addition, the developers of a growing variety of smartphone apps, including such unlikely ones as the Oxford English Dictionary (retrieved December 12, 2018, https://www.appcensus.mobi/app/com.mobisystems.msdict.embedded.wireless.oxford.conciseenglish/363), collect device location data solely in order to monetize it. A device's location may continue to be recorded even when the relevant application is not active.

The location data industry not only bridges two different sources of data but is also in a state of constant expansion and flux. There are many different actors, from telecoms giants and their subsidiaries to small independent startups, selling many different products for many different purposes. Some do not offer location data per se but instead provide customized data analytics. Among the data vendors, some specialize in a specific type of location data, while others are aggregators. They may offer a wide or narrow range of ancillary information on device activity and device owners. Social researchers thus have growing opportunities to join the many private businesses and public agencies that purchase customized location data extracts. However, only some types of available data conform with ethical standards for academic research (see the discussion of privacy and consent issues in the concluding section of this article).

The present data set. Given that our use of location data for analyzing mass protest participation is without precedent,[5] new methods of preparation and analysis had to be devised. In what follows, we introduce the data set, including the procedures used in order to prepare and then utilize it for analysis.

Our data vendor entered an agreement with Israel's largest cellular network operator at the time to continuously obtain real-time information on the geographical location of network users derived from passive monitoring.[6] The resolution of this type of location information depends on the size of the cells into which a network is divided, which in turn varies with local population density. A concrete indication in our case is that the data vendor required that the "target area" defined for protest events (which were always held in city centers) be at least 200 $\times$ 200 meters in size.

Under conditions specified by government regulators, the vendor was permitted to store the data and sell either analytics or extractions to interested clients. In the raw data automatically collected by cell networks, devices are identified by their internal hardware ID. To protect users' privacy, our vendor was required to anonymize these IDs and did not have access to the databases used by the network operator to link devices to phone numbers and to personal information on their owners.

The data processing carried out by the vendor began with imputing the residential location of phone owners by identifying recurrent sightings of devices late at night. Instead of storing the specific geo-location obtained for each device, it was coded by census tract. In Israel, census tracts (termed statistical areas) are defined as follows. All localities with a population below 10,000 are treated as a single area, while larger towns and cities are subdivided into units of 3,000–4,000 residents, which are intended to coincide with distinct neighborhoods.[7] For our purposes, working with this geographical unit had several advantages. It greatly reduced concerns about the precision of our location data, and it allowed us to merge participation data for census tracts with the many official statistics that are also aggregated at this level. (From now on, we will refer to the statistical areas to which the owners of mobile phones are linked as their *home localities*). A further benefit of utilizing locality-level data is that aggregation makes it difficult if not impossible to identify individuals.

Our first step was to establish both the times and locations (target areas) of all protest events. With this information in hand, the vendor queried their

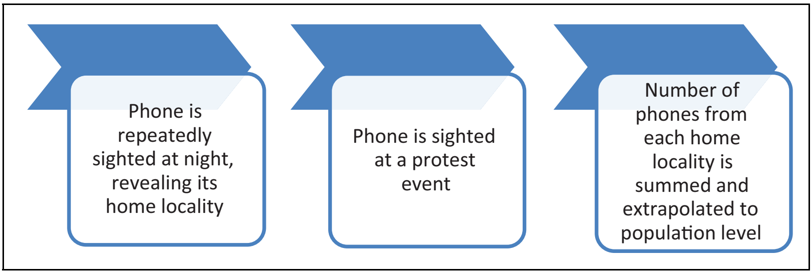

Figure 1. How geo-located phone sightings became a locality-based protest database.

database of stored sightings and computed estimates of the number of persons present at each event and their distribution across home localities. Since the source was only one cellular network, the number of devices from each home locality that was sighted at each event was extrapolated to the population of probable mobile phone owners in each locality. The vendor used a proprietary statistical model that took into account the sampling ratio, that is, the number of mobile phone users served by the network in each home locality relative to the size of its population. Finally, as a further means of ensuring anonymity, before the data set was handed over to us, all counts were rounded to the nearest 50. Figure 1 summarizes the main steps involved in creating the data set.

To clarify, for present purposes, a protest event is a demonstration that took place at a specific time and in a specific location, regardless of whether it was part of a coordinated action carried out simultaneously in other locations. Our list of 45 events was based on reports in both mainstream and social media on demonstrations and rallies held around the country between July 23 and September 10, 2011, which each reportedly included at least 1,000 participants. Usable data were extracted for 38 of these events.[8] Many small-scale local events held during the campaign did not pass this threshold. At the other extreme, mass mobilization in the 2011 campaign was concentrated mainly in a series of five coordinated protests held on Saturday nights, in all but one case with their epicenter in Tel Aviv. For each of these nationwide protests, our data set includes events that took place in as many as 18 different locations.

Adjustment for bystanders. Whether they are large or small, demonstrations usually take place in city centers where people may be present for a variety of other reasons. In addition to the residents of these areas, outsiders may arrive for the purpose of work, social interaction, shopping, or recreational activities. As a result, the number of mobile devices counted at a demonstration includes not only protesters but also *bystanders*, defined as persons present at the time and place of a demonstration who are *not there in order to join the protest*. However, as with any other method of quantifying protesters based solely on observed behavior, we lacked information on the motives of the individuals present.

We responded to the challenge of distinguishing protesters from bystanders in two complementary ways. The first was to strive for accuracy in defining the target area of each protest event, in the hope of excluding as many bystanders as possible. The time–space coordinates of the routes and sites of demonstrations were determined mainly on the basis of announcements by event organizers and mass media reports. (Systematic observations in real time and video recording of the events would have been helpful but were not available.) Other parameters that influence measurement accuracy are not under the researcher's control. One of these is the nature of each event's space. It is much easier to identify the borders of a demonstration held in a single bounded area than a march through an area in which people are living and/or other activities are taking place. Another problem is the margin of error inherent in passive location data, which as noted earlier is greater than for information collected by smartphone apps using built-in location services.

Since the demarcation of target areas cannot be perfect, and since often bystanders are mixed with the protesting crowds, we also developed indirect methods of assessing the number of bystanders at each event, broken down by their home locality. By design, our data extraction included not only counts per home locality for each protest event but also parallel "benchmark" counts computed exactly a week before. It was clear, however, that these counts would be an exaggerated proxy for the number of bystanders at the site during the protest. At least some people previously in the area for everyday reasons would refrain from returning on the day of a large demonstration in order to avoid the commotion. Among the remainder of those previously spotted at the scene, some might return in order to join the protest.

Accordingly, before subtracting week-before benchmarks from the number of persons spotted at a protest, they needed to be *deflated*. Using three quantifiable criteria, for each protest event, we computed a composite deflation rate, customized for each locality represented in the crowd. We reasoned that a benchmark count would be less likely to represent bystanders (and hence more in need of deflation) if (a)residents of the home locality were often seen at other protest events during the campaign, (b)the event was

Table 1. Estimated Crowd Size at the Two Largest Protest Events (Thousands).

	Location Data	Media Reports
Tel Aviv, August 6	219 (218.1–219.8)	230–300[a]
Tel Aviv, September 3	123 (119.1–127.1)	125–280[b]

Note: In the location data column, figures in parentheses are the range of estimates derived from multiple schemes of adjustment for bystanders (see the text and Online Appendix 1 [which can be found at http://smr.sagepub.com/supplemental/] for details).
[a]*Source*: net.nana10.co.il/Article/?ArticleID=820449 (accessed October 19, 2011).
[b]*Source*: it.themarker.com/tmit/article/16771 (accessed October 19, 2011).

located relatively far from home, and (c)the event location was unlikely to attract everyday visitors.

The specifics of how we went about distinguishing presumed protesters from presumed bystanders, including deflating the benchmark counts, are provided in Online Appendix 1 (which can be found at http://smr.sagepub.com/supplemental/).[9] In addition, we tested the sensitivity of a wide range of our findings to different ways of computing the composite deflation rate. The results, reported in Online Appendix 2 (which can be found at http://smr.sagepub.com/supplemental/), show that this made very little difference to the findings. Our estimates of the size of protest events are similar in order of magnitude to those reported by the media at the time of the protest but more conservative. Table 1 compares the estimates for the two largest protest events (both held in Tel Aviv). It also includes the lower and upper bounds of our estimation derived from different ways of computing the composite deflation rate. The table shows that our figures lie near the lower end of the ranges reported by the media. Note the narrow range of measurable uncertainty introduced by the procedure for adjusting for bystanders.

To reiterate, our estimated protester counts (for each event and home locality) are the gross number of persons observed in the event's target area, less than the deflated benchmark number from a week before. Given that demonstrations can last for a number of hours, researchers must decide how frequently counts should be made and which of them will serve as the count(s) of record. In the data set acquired for our research, headcounts for each event were supplied for the beginning of each hour. We selected the highest of these hourly counts since it is closest to the actual number of persons who participated. If data on individual phones over the entire course of each event had been available, it would have been possible to more accurately estimate the total number of individuals present, irrespective of

Table 2. The Data Set Matrix.

	Protest Events			
	Event 1	Event 2	Event n	Total Participations per Locality
Home localities				
Locality 1				
Locality 2				
Locality n				
Total participations per event.				

when they arrived and left. Since this is not the case, our participation estimates are biased downward.

The Data Set, How It Can Be Analyzed, and the Measures It Yields

The procedures described above were followed to generate estimates of the number of participants from each home locality present in the 38 events included in our data set. As mentioned, home localities are the census tracts where protesters live. To minimize reliability issues, our working data set is limited to tracts with at least 300 working-age residents (aged 18–64, approximating the number of potential protesters). A total of 2,237 home localities for which we obtained estimates of protest participation passed this threshold, comprising 96 percent of the working-age population of Israel.[10] Both the mean and the median size of the included home localities are approximately 1,800.

In the working data set that is the basis for our empirical analyses, estimates of the number of protesters (after deducting bystanders) are arrayed in a matrix consisting of protesters' home localities in the rows and protest events in the columns. Table 2 shows the structure of this matrix, with the addition of marginal totals shaded in gray.

The *row totals* in Table 2 offer a *campaign-oriented perspective*. They constitute the total number of "participations" by residents of each locality at all of the events in the protest campaign.[11] By coding other home locality attributes, for example, their ethnic or educational composition, it is possible to explore what kinds of localities were highly mobilized in the protest compared with those which played little or no role.

We generally measure home locality mobilization by converting each row's participation *totals* to participation *rates*, relative to the size of the potential protesting population. Associations can then be explored between these rates and other locality attributes. For this purpose, numerous attributes available at the level of statistical areas from Israel's Central Bureau of Statistics and other public agencies were merged with our measures of protest participation rate. These included data on the socioeconomic level, ethnic and religious composition, and voting behavior of the residents of each home locality.

The *column totals* in Table 2, which indicate the total number of protesters at each event, enable an *event-oriented perspective* on the data. Contrasts can be made between participation at specific events or in reference to event characteristics such as when they occurred (different phases of the campaign) or where they took place (e.g., in large cities versus more peripheral locations).

A third and more nuanced way of exploiting the data matrix is to combine information from its rows and columns. From an event-oriented perspective, it is possible to explore whether different types of events attracted different crowds (based on the attributes of the home localities from which they came). Correspondingly, when analyzing the drivers of variation in home locality mobilization, the protest campaign may be disaggregated at the level either of individual events or different types of events. This opens the way to discovering variation in the profile of participants at different events or changes over the course of a campaign in the social bases of mobilization.

The appropriate way of counting participations depends on which of the two analytical perspectives is adopted. From a campaign-oriented perspective, we considered two global measures of home locality mobilization. The first of these, *total participations*, aggregates the counts obtained for each locality at all events. The second measure, *peak participation*, is based solely on counts for the specific event to which each locality sent the highest number of demonstrators.

While total participations is the more comprehensive of these two measures, it is often preferable to focus on a locality's best effort (its "peak participation"). The share of the members of a community who mobilized cannot be calculated using total participations, which include an unknown number of repeat demonstrators. The same home locality total could be generated by a variety of scenarios, from one-time attendance by different members of the community to repeated participation by a smaller group of dedicated activists. Accordingly, the ratio of peak participation to the size of the working-age population is our preferred measure of localities' engagement in the campaign as a whole.[12]

Table 3. Three Methods of Collecting Data on Participants in Large-Scale Protests.

	Population Surveys	On-site surveys	Location Data
Unit of analysis	Individuals	Individuals	Localities
Time of data collection	After the event	In real time but requires preparation	Gathered automatically in real time, available after the event
Sample	Small systematic sample	Small systematic sample	Large "found" sample
Coverage	Depends on sampling frame and respondent cooperation	Depends on conditions. Nonparticipants not covered	Depends on cellphone penetration and network coverage
Time and space resolution	Not applicable	Not applicable	High
Source of participation data	Self-reports	Behavior	Behavior
Source of data on covariates	Self-reports	Self-reports	Census, administrative and event data

The alternative event-oriented perspective focuses on spatial and temporal patterns of mobilization. It is possible to compare the composition of participants at different event locations, events held at different times during the campaign, or a combination of the two. For example, the dynamics of a campaign can be tracked over different phases by analyzing changes in the makeup of the crowds, namely the distribution of participants by their home localities. As in analyses from a campaign-oriented perspective, most event-oriented results are presented as relative rates rather than absolute numbers.

Comparing Methods of Data Collection

Now that our data and methodology have been introduced, and before moving forward to give examples of its uses, Table 3 summarizes the differences between using automatically collected mobile location data to study mobilization in large-scale protests and the most viable alternative sources—either population-wide or on-site sample surveys. Because both types of survey collect information directly from individual respondents, they offer researchers the opportunity to try to connect engagement in protest to personal backgrounds, outlooks, and a host of other theoretically grounded causal variables such as sense of grievance, sympathy for the protest, prior protest experience, personal contact with activists and supporters, and availability for participation (e.g., Anduiza, Cristancho, and Sabucedo 2014; Rüdig and Karyotis 2014). On the other hand, since even on-site surveys suffer from potentially severe time inconsistency problems, the information they provide (including on motives for joining a protest) is not necessarily reliable.

The superior size and coverage of the samples available using cellular big data, as well as the fact that the crowd counts it yields are based on observed behavior monitored in real time, offer substantial benefits for researchers interested in mapping the social, spatial, and political coalitions underpinning mass protests. Moreover, as will shortly be shown, the use of localities as the unit of analysis affords novel research opportunities that are especially relevant to protest campaigns comprising events at multiple sites and times.

Clearly, as with all methodological choices, trade-offs are involved. While the type of behavioral data we use is free of the potential distortions inherent in survey samples and self-reports, it is subject to several other types of error. These include inaccuracies in location information and imperfect population estimates when using data that originate from a single mobile carrier, as well as uncertainty regarding the number of bystanders whose presence in the area of protest events is recorded along with that of authentic protesters. Yet these caveats are more likely to affect estimates of the absolute number of participants than either their composition or their time–space dispersion over the course of a protest campaign, features that we argue are of greater importance to researchers.

Other issues arise because of the aggregated nature of the data on protesters when seeking to profile the protesting and nonprotesting populations, measure group differences in mobilization, and isolate the net effect of specific attributes of potential protesters on their likelihood of participation. No difficulty would arise in making such inferences from spatially aggregated data if the residents of home localities were always homogenous with respect to either the outcome of interest (joining the protest or abstaining) or its presumed determinants (e.g., if all belong to the same ethnic group, vote for the same party). To the extent that localities are internally heterogeneous, researchers run the risk of making erroneous inferences from geographically aggregated data (the ecological fallacy).

This and other potential limitations of location data are discussed in the concluding section of this article. Our task now is to demonstrate how a data set of this type is constructed and can be used for research, using the case of a mass protest campaign in Israel in 2011.

Part 2: Applications to Israel's 2011 Social Justice Protest

Between July and September 2011, Israel was swept by a protest that mobilized hundreds of thousands of citizens and completely dominated public discourse (Ram and Filc 2016; Rosenhek and Shalev 2014). It began in Tel Aviv as an outcry against the scarcity and cost of rental housing in the city and quickly expanded in scale, geographical spread, and demands. The protest articulated widespread discontent on a variety of social and economic issues, bundled together by a critique of neoliberal policies and an overarching demand for social justice. Its leaders and activists adopted inclusive rhetoric, refused to align with or take positions on divisive noneconomic issues, and made explicit efforts to draw a variety of social groups into the ranks of the demonstrators.

The social justice protest quickly came to enjoy widespread support,[13] obliging even the political leaders whose policies were under attack to recognize the public's distress and promise to relieve it. Except at times of war and conflict, such a broad consensus among both elites and the mass public is rare in the Israeli context, in which long-established cleavages between rival political camps and identity groups are rarely challenged. Nevertheless, the centrality of the plight of the younger generation of the middle classes to the discourse of the protest, and the predominance among its leaders and most visible supporters of secular and college-educated people that oppose the dominant right-nationalist political bloc, risked alienating social and political sectors outside of these orbits. As a result, Israel is an ideal setting for illustrating the boundaries of recent protest campaigns that shunned conventional politics and spoke in the name of "the people." As noted in this article's introduction, the question is not only who demonstrated and who stayed home but also who protested with whom.

The analyses summarily presented in this section of this article aim to illustrate a range of novel uses of the data set introduced in part 1. More theoretically driven questions and more elaborate types of data analysis will be taken up in other papers. In line with our aim of encouraging social movement researchers to take advantage of mobile location data, the concluding section (part 3) includes a typology that summarizes the variety of causal effects that are testable using geographically aggregated data on protest participation, including topics not empirically addressed here.

In what follows, we begin by exploring the campaign as a whole, first to gauge fundamental aspects of variability in protest participation rates and then to gain insight into the social and political backgrounds of the protesters. After that we turn our attention to temporal and spatial variation over the course of the campaign. By looking at differences in the composition as well as intensity of mobilization, the findings offer suggestive evidence that different phases of the campaign and different sites of protest varied not only in the size of the crowds they attracted but also in who turned out.

The third and last set of empirical illustrations shifts the focus from the correlates and dynamics of mobilization to its role as an independent variable. Specifically, we ask whether the data provide at least prima facie evidence that the extent to which a community is engaged in protest influenced other outcomes of interest to scholars. One of the examples, which is based on survey data merged with our data set, indicates that the aggregate mobilization of a local community has a contextual effect on the probability of participation by its individual members. The second example shows the downstream effect of the intensity of mobilization in home localities on the first elections held after the protest.

The Variability of Participation Rates and the Diversity of Participants

One way to assess the breadth of the mobilization that occurred in Israel during the social justice protest is by examining the geographical dispersion of both the *sites* of protest and the origin of *protesters*. The maps in Online Appendix 3 (which can be found at http://smr.sagepub.com/supplemental/) attest to a high degree of spatial concentration in both of these respects. While demonstrations were held throughout the country, Tel Aviv and several lesser epicenters drew by far the largest crowds. Similarly, the vast majority of all participants were residents of the most densely populated regions in the center of Israel (the Tel Aviv and Jerusalem metropolitan areas). Nevertheless, high participation *rates* were much more dispersed across home localities around the country. This is explained both by the opportunities provided by local events and the fact that highly motivated participants were ready to travel in order to take part in major demonstrations.

Sectoral cleavages and protest mobilization. Variation in protest participation can also be assessed quantitatively. We analyze home locality mobilization using a two-step approach, asking first whether localities passed some minimal threshold of protest activity and then addressing variation in the participation rates of active localities. Similar to the analysis of individual engagement in protest, each of these two might be explained by different factors (Klandermans 1997:chapter 1).

We classified localities with a participation rate[14] of less than 1 percent of their working-age population as "inactive" and those exceeding 1 percent as "active." About 60 percent of the total of 2,232 home localities with nonmissing data passed this minimal threshold. Among these active home localities, the level of participation varied widely (the coefficient of variation is 1.26).

In the context of the Israeli protest, the main interest in protester composition revolves around the tension between the broad backing it enjoyed from nearly all social sectors (supported by the inclusive rhetoric invoked by protest leaders) versus public perceptions and survey evidence suggesting that the leaders and core activists were drawn primarily from Israel's historically hegemonic sector, comprised of relatively affluent, educated, secular, and left-leaning Jews. A strong test of the diversity of those who were mobilized is the level of engagement of sectors of society that either did not support the protest or else faced clear cultural, political, or geographical barriers to translating their sympathy into action.

We focus on three minority groups that media reports and survey findings indicate fall into this category: Palestinian Arab citizens, ultra-orthodox Jews, and "settlers," the residents of Jewish settlements in the Palestinian territories known as the West Bank. Collectively these three sectors comprise one third of all Israeli citizens. Arabs are the largest minority (18 percent), followed by ultra-orthodox Jews (11 percent) and settlers (5 percent).[15] Substantively, there are grave tensions between all three of these minorities and the secular Jewish middle class, and they seldom cooperate in the political sphere.[16]

By definition, settlers live in a clearly demarcated geographical area. The Arab and ultra-orthodox populations also live almost exclusively in homogeneous and spatially segregated residential communities. Consequently, home localities are a reliable basis for distinguishing all three of these sectors and comparing them to the residual "majority."

Figure 2 presents the two facets of locality-level mobilization: the proportion of "active" home localities (those that crossed the minimal 1 percent threshold) and, for active localities only, their average participation rate. As expected, both the activity rate and the participation rate of the minorities are far below those of the majority. Weighting the results by the size of

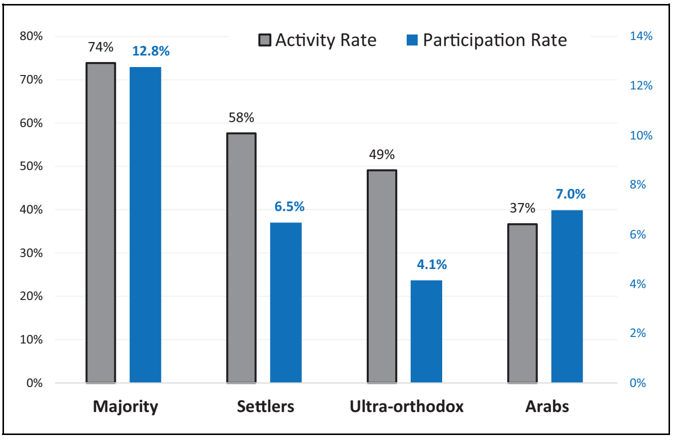

Figure 2. Mean rates of activity and participation, by population sector. Population-weighted means of home localities. "Active" localities are those with a peak participation rate of at least 1 percent. Participation rates are for active localities only.

localities' working-age population, we estimate that 37 percent of Arabs, 53 percent of Jewish ultra-orthodox, and 57 percent of settlers live in localities that were at least minimally active, compared with 74 percent of the rest of the population. The average level of mobilization in active home localities populated by the majority is roughly twice as high as in two of the minority sectors, with ultra-orthodox localities lagging even further behind.[17] Another interesting finding is that while nearly two thirds of Palestinian Arab citizens live in localities that took no meaningful part in the protest, participation in those that did cross the minimum threshold is the highest of the three minorities.

The partial participation of seemingly alienated minorities supports the claim that the protest had at least some success in reaching out to diverse groups including those with distinctive identities, worldviews, and political preferences. At the same time, considerable internal variance in mobilization is evident among both the majority and the minorities. It would certainly be possible to delve into the circumstances that promoted or hampered the involvement of each of the populations least likely to actively join the campaign, but for present purposes, the important findings concern the sheer dimensions of their participation. Our behavioral data provide more credible indications on this score than population surveys, the only other source of evidence that is more than impressionistic. The findings of a systematic comparison reported in Online Appendix 5 (which can be found at http://smr.sagepub.com/supplemental/) suggest that two unrelated population surveys significantly overestimated minority participation. We attribute this mainly to pollsters' reliance on small minority samples that fail to accurately represent these "hard-to-reach" populations. Besides the effect of group identity, other drivers of mobilization were also at work, and their influence will now be assessed alongside sectoral effects.

Explaining variation in local mobilization. We chose three measurable determinants of variation in home localities' protest engagement: (1) socioeconomic differences (level of education), (2) geographical differences (accessibility of protest sites), and (3) political differences (vote distribution).

1. *Socioeconomic differences:* It is a truism that in contemporary societies, the distinction between the college-educated and those without higher education is fundamental for both life chances and social status. Therefore, the share of college graduates in localities' populations serves as a reliable measure of the socioeconomic status of their residents. Furthermore, since the leaders of the social justice protest and most of its activists were either students or young professionals, and one of its central themes was the declining opportunities and living standards of these two groups, the college/noncollege binary is a promising predictor of mobilization. The necessary data are robust and readily available at the locality level in the form of the proportion of working-age adults with college degrees.
2. *Geographical differences:* While personal availability for protest due to circumstances like age and family responsibilities has long been recognized as a determinant of individual participation, the opportunity to access a protest event (a function of distance) has received much less attention (but see Converse and Pierce 1986; Rüdig and Karyotis 2014; Traag, Quax, and Sloot 2017).[18] Locality-level data are ideally suited for this purpose. The distance between each home locality and the nearest event (host locality) was obtained from a GIS matrix.
3. *Political differences:* As in other recent encompassing and ostensibly apolitical mass campaigns, there is evidence from surveys that political cleavages played a decisive role in structuring participation in Israel's social justice protest (Perugorría, Shalev, and Tejerina 2016). However, since party competition and voting in Israel are unusually

detached from conflicts over social and economic policy (Shalev 2007; Shamir and Arian 1999), a brief explanation is required.

Ever since the historic 1967 war, the political left and right have primarily been divided by issues related to the self-image of the state (Jewish vs. democratic) and Israel's military posture and territorial aspirations (hawk vs. dove; Shamir, Dvir-Gvirzman, and Ventura 2017). While these contested issues did not bear directly on the 2011 protest, and indeed were intentionally sidestepped by its leaders and most activists, the left–right division was salient for other reasons. The main political camps represent and promote fiercely competing worldviews and collective identities. These usually preempt participation in collective action under alien sponsorship, even in cases where common material interests are at stake. Moreover, because the protest was directed against the policies of the incumbent government, and might therefore have threatened the short and even long-term dominance of the ruling Likud party and its nationalist and religious coalition partners, their supporters also had a pragmatic reason to stay away.

Among the country's Jewish majority, supporters of right and religious parties far outnumber leftists, but between them is an often sizable center bloc that on many issues is closer to the left than the right. Accordingly, a simple but potentially powerful predictor of the extent to which active home localities joined the protest is the combined vote share of left and center parties in the 2009 election, the last one before the protest.[19]

Bivariate rank correlations between the above three determinants and the two facets of local mobilization (whether localities are active and if so how intensely) confirm that the influence of different drivers depends on the outcome. The strongest correlate of activity is distance from the nearest event (Spearman's $\rho = -.35$), while the variable most strongly associated with the participation rate of active localities is the vote share of left and center parties ($\rho = .45$).

To investigate the relative strength and significance of the independent variables, we performed a two-phase multivariate analysis. First, a logistic regression was used to predict the net probability of activity (participation of at least 1 percent). Next, for active home localities only, we used a GLM model to predict net differences in their participation rate. To deal with nonlinearity, both estimations are based on multiple fractional polynomial models.[20]

Figures 3 and 4 that follow provide visualizations of the most important findings from these regressions. Note that the results of these (and subsequent) models are reported in Online Appendix 6 (which can be found at

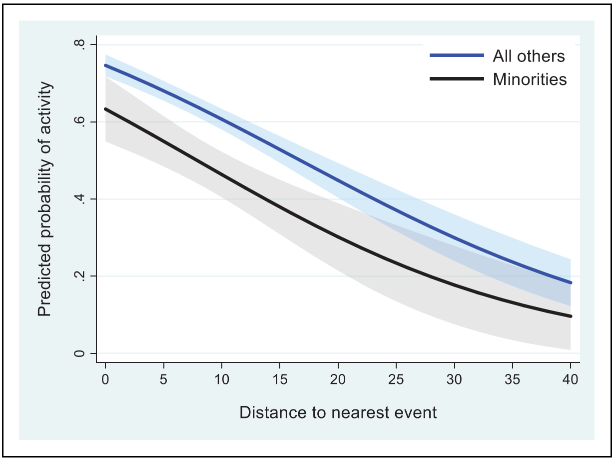

Figure 3. Effect of event accessibility on the likelihood of minimal protest activity, by population sector. Margins plot of the net effect of home localities' distance from the nearest protest event in kilometers on their probability of being active, defined as a participation rate of at least 1 percent of the working-age population. See the text for model specification and Online Appendix 6 (which can be found at http://smr.sagepub.com/supplemental/) for results.

http://smr.sagepub.com/supplemental/), along with the number of home localities included in each analysis.

The most powerful predictor of "activity" is the geographical distance of a locality from the nearest protest event. Figure 3 plots this relationship, conditional on group membership (a binary variable contrasting the three minorities collectively with the majority). Note the persistent gap between the two lines, indicating that even after the effects of the control variables are accounted for, minority localities have a lower probability of activity.[21] The slope of the lines indicates that geographical proximity greatly facilitates mobilization by both groups. However, and not surprisingly, in proportional terms, the chilling effect of distance is greater for minority than majority localities. Compared with home localities in cities that hosted an event, those 10 kilometers away from the nearest protest are 21 percent (majority) and 31 percent (minority) less likely to pass the active threshold. This effect intensifies as distance increases.[22]

Next, we focus solely on active home localities, examining associations between their logged participation rate[23] and the same set of determinants; however, this time the minorities are modeled separately. The strongest

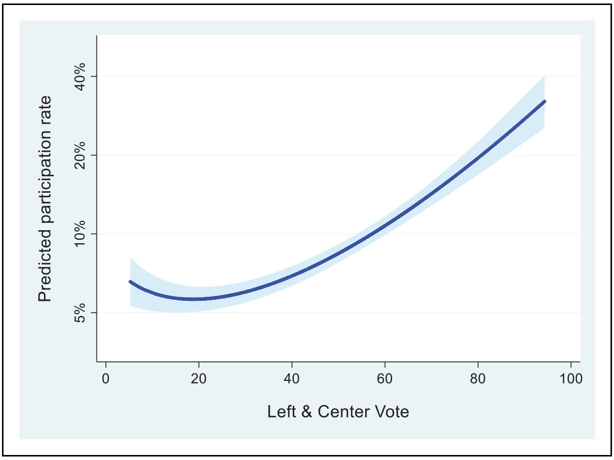

Figure 4. Relationship between protest participation and support for left and center parties. Fractional polynomial plot of the net effect of the vote share of left and center parties in the 2009 elections on home localities' participation rates in the 2011 protest. *Y*-axis uses a log scale (base 2). Analysis includes only home localities that were "active" (participation rate of at least 1 percent) and had a left and center vote share of at least 5 percent. See the text for model specification and Online Appendix 6 (which can be found at http://smr.sagepub.com/supplemental/) for results.

predictor of locality participation is the left and center vote share, and Figure 4 plots this relationship. Mobilization is higher in places where a larger proportion of the population voted for these parties in the preceding elections. This effect is strong but not linear, mainly because it emerges only after the share of left and center supporters rises beyond a threshold of about 30 percent.

In addition to the powerful effect on mobilization of the political division between the main electoral blocs, model results show that localities' socioeconomic status and their distance from the nearest event have weaker but statistically significant effects. Although as expected all three dummy variables representing the minority sectors have negative coefficients, only the one for ultra-orthodox Jews is significant.

To sum up this exploration of the determinants of home locality mobilization over the campaign as a whole, both the descriptive findings and the results of the multivariate models point to conclusions that are plausible and meaningful. *First*, they confirm that populations that are socially and politically remote from the main backbone of the Israeli protest indeed joined the demonstrations but at substantially lower rates than had been reported by

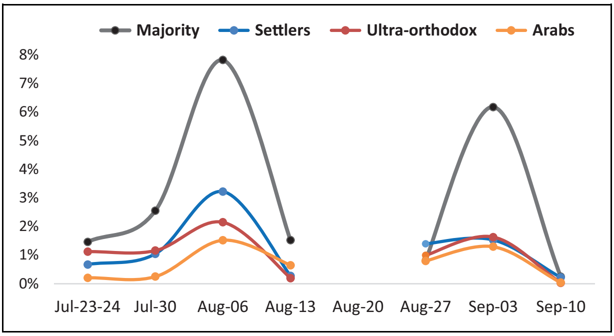

Figure 5. Protest participation over time, by population sector. Participation rates are the share of a sector's working-age population participating in the protests held on each date.

population surveys. *Second*, the finding that spatial accessibility is the major determinant of whether more than a trivial share of a locality's residents participated in the protest draws attention to a fundamental barrier to participation that has rarely been studied. Using our geo-tagged behavioral data, the distance effect could be easily determined and also reliably modeled given the very large sample of mobile phone users whose location was monitored. *Third*, consistent with earlier findings based on individual-level surveys, political partisanship plays a paramount role in explaining variations in participation between active home localities.

Temporal Dynamics and Spatial Dispersion

Because the data matrix used in this research provides information on who protested (based on their home localities) for each protest event, it is possible to go beyond inferring the role of different drivers of participation. The findings that will now be presented address the questions of *when*, *where*, and *with whom* different groups joined the campaign.

Compositional changes in the course of the campaign. We have already seen evidence of substantial differences in aggregate mobilization between social sectors. Now, we disaggregate the intracampaign dynamics of participation, focusing on groups already shown to have been unlikely protesters. To the extent that these groups participated at all, were they drawn into the ebbs and flows of mobilization on the part of more likely protesters or was the timing of their participation idiosyncratic? Building on earlier findings, we first

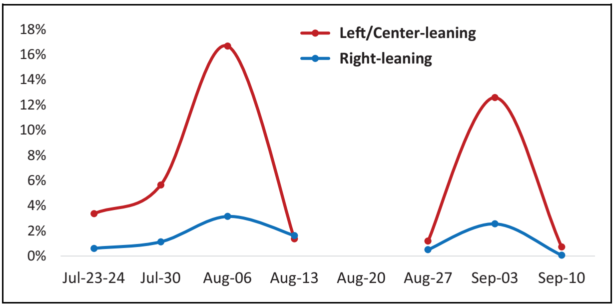

Figure 6. Protest participation over time, by political partisanship. Excludes the three minority sectors. Lines differentiate between home localities in the top and bottom tertiles of the vote share of left and center parties in the 2009 elections. They report the share of the working-age population which participated in the protests held on each date.

address the distinction between the three minority sectors and the residual majority. Then, within the majority, we look at the political contrast between localities that strongly support left and center parties and those where parties of the right are strongly supported.

Sectorally and politically disaggregated trends in the evolution of mobilization are presented in Figures 5 and 6. The trend lines are interrupted by a weekend in which the protesters took a time out due to a terror attack on Israel's southern border. This break visually underscores the main feature of the campaign's temporality: two distinct cycles in which participation rates rose sharply to a peak and then fell off again. Despite large intergroup differentials in protest *intensity*, clear similarities are evident in the *trends*. In particular, both Figures 5 and 6 show that in the run-up to each peak, a rising tide lifted nearly all boats.[24]

Based on the data underlying Figure 6, the ratio of participation in left/center-leaning versus right-leaning localities is approximately 5:1 in the first two large-scale mobilizations and at both of the peaks. However on August 13, when mobilization plummeted following its first peak, the participation rate of right-leaning localities was actually a little higher than for left/center-leaning localities. This was the intended result of a top-down initiative, assisted by local activists, to temporarily suspend protests in the two epicenters in the center of the country (Tel Aviv and Jerusalem) in order to spotlight and include outlying areas and increase the engagement of sectors outside the main group of protesters (mainly the Jewish right-leaning working class). However, the call for mobilization of "the periphery" was not directed at the three minority sectors, and this inclusionary moment of the campaign bypassed all three of them (Figure 5).

These findings show that disaggregated temporal data on mobilization over the course of a protest campaign can shed new light on how movements that strive for encompassingness succeed in incorporating sectors beyond their natural constituency. The similar trends in the mobilization of diverse groups as the Israeli protest moved toward both of its two peaks are indicative of its success in incorporating those less likely to join the crowds, albeit at much lower intensity than core supporters. The findings also suggest that at different moments during a protest campaign, the power relations between groups of protesters can change: The modest but significant success of the events held on a date explicitly reserved for participation by politically alienated supporters suggests that separation can play a role in recruiting reluctant supporters to a campaign with inclusive ambitions. This suggestion receives additional support in the next section, which reports findings related to the spatial dimension of intracampaign variability.

Spatial variation in protest participation. In large-scale protest campaigns that seek to engage diverse sectors of society, and which offer them a variety of locations at which to demonstrate, several forces may influence protesters' choice of venue (AlSayyad and Guvenc 2015; Braun and Koopmans 2010; Myers 2010). These include (1) convenience, that is, the accessibility of available event sites to potential protesters; (2) the aura surrounding "magnet" protest sites, like Tahrir Square in the Egyptian revolution; and (3) homophily, the tendency for people to seek the company of others like themselves. It can further be assumed that the importance of these three considerations varies between different groups. For example, we expected that convenience would weigh more heavily in the decisions of protesters who are less committed but might be overridden by the power of attraction of mass gatherings held at the central site of protest. We also anticipated that the motivation to congregate with other socially similar protesters would be especially strong for ethnically or culturally distinct sectors, particularly if their customs limit face-to-face interactions with outsiders.

The strength of the convenience effect, for both minority and majority social sectors, has already been shown in Figure 3. However, further analysis reveals a striking difference between Tel Aviv and other event locations (see Online Appendix 7, which can be found at http://smr.sagepub.com/supplemental/). The mega-demonstrations held in Tel Aviv exerted such a magnetic attraction that distance has no chilling effect on participation by protesters living outside the city's metropolitan area.

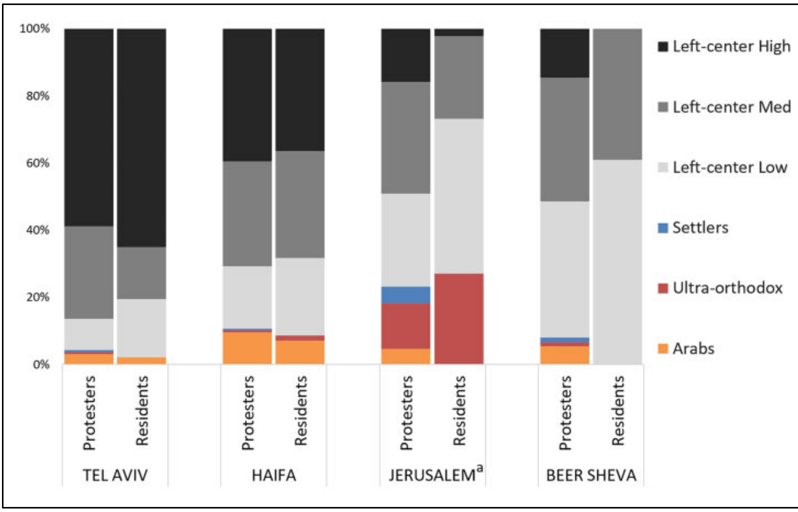

Figure 7. Composition of out-of-town protesters and local residents at major event locations. Protester composition is based on the total number of participations at each event location, local residents excluded. Left and center support level refers to the "majority" sector and is measured by tertiles of the vote shares of left and center parties in the 2009 elections. [a]The local population in Jerusalem does not include the Palestinian neighborhoods of East Jerusalem.

Homophily, a third hypothesized influence on the spatial distribution of protest, is tested by comparing the composition of the crowds that gathered in different locations (with local residents set aside because in their case it is more difficult to distinguish between participants and bystanders). A simple but telling illustration is presented in Figure 7, which for each of the campaign's four leading event locations compares the characteristics of out-of-town protesters with those of local residents. Similar to the overtime analysis, the three alienated minorities are distinguished, and the remaining localities are subdivided by tertiles of support for left and center parties.

Figure 7 confirms our expectation that the protesters from outside the host city who assembled at events in each of the four epicenters differed substantially in their composition. Furthermore, the composition of protesters is generally very similar to that of the local population of the host cities. Reflecting the character of the city, participants in Tel Aviv were less inclusive of minority sectors and more politically homogenous. Similarly, the almost exclusive role of Jerusalem in hosting ultra-orthodox protesters and the above-average representation of Arab citizens at events held in Haifa suggest that minority groups indeed prefer to participate in places that are not only more physically accessible but also more heavily populated by people like themselves. Striking variations are also evident in the presence of participants from left-leaning and right-leaning localities at different sites of mobilization. Consistent with the homophily hypothesis, Figure 7 makes evident the similarity between the political profile of local residents and out-of-town protesters. Looking across all 17 host localities, we found a strong bivariate association between support for left and center parties in host and home localities (Spearman's $\rho = .73$, $p < .001$).

These findings underline the significance of a mechanism that in Israel clearly facilitated encompassing collective action—namely, temporal and spatial segmentation of protesters representing groups with conflicting identities. Simply put, collaboration between social groups in the course of a protest campaign does not have to mean that protesters from different sectors literally stand shoulder to shoulder. In the words of Walgrave and Verhulst, whose pathbreaking work on this topic has not yet received the attention it deserves, it is entirely possible that "different people take to the streets but not to the same streets at the same time" (Walgrave and Verhulst 2009:1357). Indeed, some degree of time and place segmentation may even be a precondition for encompassing collective action that succeeds in mobilizing diverse publics, especially in heterogeneous and divided societies. The high geographical and temporal resolution of location data, combined with sample sizes that easily permit investigation of protester heterogeneity, suggests that this is one of its most promising uses.

Effects of Protest Participation

The final empirical illustrations reverse the direction of the causal arrow, showing how data on locality-level protest mobilization may be utilized as an independent variable in order to explain both the dynamics and effects of a mass protest campaign. The first of these novel applications is based on merging our data on the aggregated protest participation of residential localities with individual-level data on their residents. A multilevel data set like this has many possible uses, two of which are illustrated here.

Contextual effects of local mobilization. Political action is rarely taken in solitude, and it has long been recognized that rather than focusing solely on individual-level characteristics, the local social and political context should also be considered as a factor influencing participation. A seminal study by Huckfeldt (1979) suggested that when political participation is the norm in a

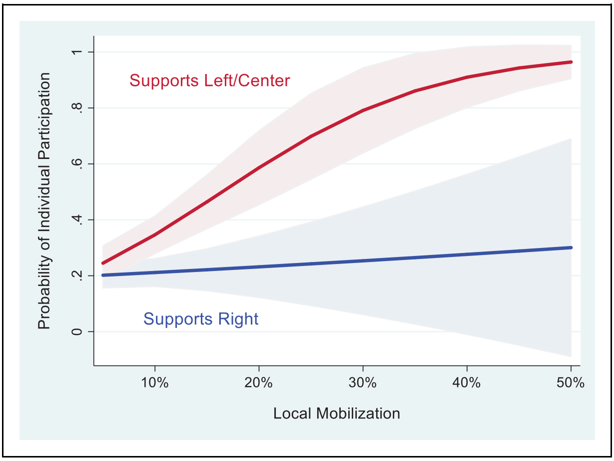

Figure 8. Individual participation rate by individual political ideology and level of local mobilization. Estimates derived from a multilevel logit regression. Local mobilization is the home locality peak participation rate. Individual-level data are from the Israel Democracy Institute's 2012 "Democracy Index" survey ($n = 659$), based on self-reports (Jewish localities only). The model controls for individuals' age, gender, education, religiosity, and immigrant status. See text for model specification and Online Appendix 6 (which can be found at http://smr.sagepub.com/supplemental/) for results.

community and it is positively valued in one's immediate social environment, the legitimacy to act is high and participation is more probable. However, since different types of individuals are affected to different degrees by their environment, interactions are likely between individual- and contextual-level attributes. Merging our aggregated data on protest participation per home locality with geo-located data on individuals from a population survey makes it possible to explore the contextual effect of local mobilization on the protest participation of individuals, controlling for relevant background variables.[25]

For hypotheses that posit contextual effects on individuals, hierarchical models are recommended (Gelman et al. 2007). We performed a multilevel logistic regression to test the effect of the local participation rate on the participation of individuals, conditional on their a priori likelihood of participating. The latter was modeled by control variables representing predictors previously identified in survey research on the Israeli protest: age, gender, college education, political ideology, religiosity, and origin in the former Soviet Union (Haber, Heller, and Hermann 2011). Political ideology, the strongest of these predictors, was interacted with local mobilization in order to test whether the impact of community-level participation on individuals is moderated by the degree to which they are predisposed to participate.

The most important findings from the regression model are summarized in Figure 8 (full results are reported in Online Appendix 6, which can be found at http://smr.sagepub.com/supplemental/). Figure 8 portrays the relationship between local mobilization (horizontal axis) and the probability of protest participation by individuals (vertical axis), segmented by survey respondents' left-right placement. It is evident that local mobilization has no significant effect on the probability that rightwing individuals join the protest. But for those identifying with the left or center of the political spectrum, community-level engagement is associated with a dramatically higher individual propensity to demonstrate. The fact that no contextual effect was detected for individuals identified with the right lays to rest any suspicion that the correlation between individual and local mobilization is simply due to the success of the survey in sampling individuals who are representative of their community.[26] Accordingly, these results vindicate the case for further research using integrated macro and micro protest data to address the argument that the political participation of individuals is affected by norms of participation in their immediate environment.

Protest and electoral behavior. Data on local rates of engagement in a protest can also serve researchers interested in linkages between contentious politics and electoral behavior. There are multiple ways in which protests can influence elections, including consolidation of protest movements into new political parties; changes in political attitudes that shift voter preferences and stimulate changes in party platforms; and altered levels of political interest and efficacy, reflected in voter turnout. While interest in these and other interplays between "ballots and barricades" has increased in recent years (McAdam and Tarrow 2010), due to a lack of suitable data empirical studies of the impact of mass protests on election outcomes are rare (but see Anduiza, Martin, and Mateos 2013; Gillion and Soule 2018).

To explore the potential contribution of locality-level data on protest participation, we model the effect of local mobilization on aggregate local vote shifts in the elections held before and after the protest (2009 and 2013). Commentators have suggested that the protest shaped the issue agenda of the 2013 elections, with two palpable results (Shamir 2015; Talshir 2015). First, it lent new relevance to declining leftwing parties, which gained 5 more seats in the 120-seat Israeli legislature. In addition, the protest served as a springboard for a new contender—*Yesh Atid* ("there is a future")—

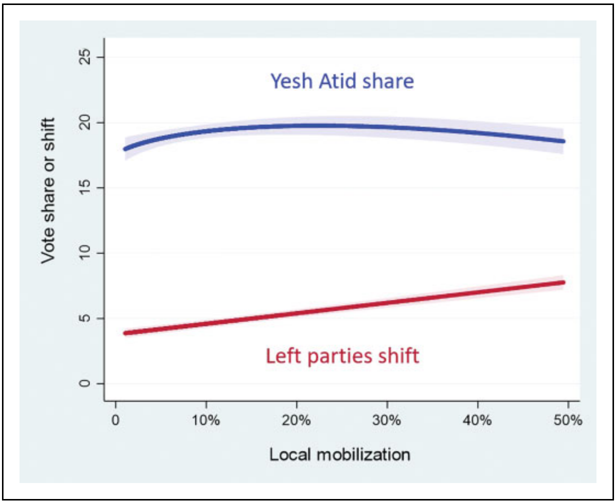

Figure 9. Electoral outcomes in 2013, by local mobilization in 2011. Chart combines estimates based on separate fractional polynomial regressions of the net effect of home localities' peak participation rate in the 2011 protests on two electoral outcomes: the share of votes cast for Yesh Atid in 2013 and the absolute percentage point change in the vote share of leftwing parties between the 2009 and 2013 elections. Minority sectors (Arabs, ultra-orthodox Jews and settlers) excluded. See text for model specification and Online Appendix 6 (which can be found at http://smr.sagepub.com/supplemental/) for results.

which coopted some key elements of its discourse. Yesh Atid garnered 14 percent of the popular vote which translated into 19 seats, second only to the ruling Likud party.

Figure 9 plots the relationship between the intensity of home locality participation in the 2011 protest and the above two local electoral outcomes (the difference between the vote shares of leftwing parties in 2009 and 2013 and the share obtained by the Yesh Atid party in its electoral debut in 2013). These results are based on separate fractional polynomial regressions that estimate the expected effect of local mobilization on the subsequent electoral behavior observed in home localities, controlling for their socioeconomic standing (approximated by the share of college graduates in the adult population). The findings for home localities with peak participation rates under about 20 percent indicate that both Yesh Atid and the left may have benefited from local mobilization. However, among localities with higher levels of protest participation, the impact on the two parties diverges. The effect on

Yesh Atid support flips, while the success of the left parties continues to rise. It is likely that participants in highly mobilized localities were attracted by the progressive agenda promoted by the protest, which was central to the election rhetoric of the left parties but not Yesh Atid, which presented itself as the party of the middle class.

These associations support our suggestion that data on protest participation at the level of towns and neighborhoods potentially have much to offer scholars interested in how protests impact electoral outcomes. Future analysis will of course need to go beyond the simplified model tested here, which cannot rule out alternative interpretations of the findings. It is possible that previously formed preferences were responsible for both protest participation and voting or that newly convinced left party and Yesh Atid voters were not necessarily the same individuals who joined the protest.

Part 3: Conclusions

Because close to entire populations carry their phones with them wherever they go, and the location of these devices is automatically recorded and stored, a new frontier has opened for measuring and analyzing participation in large-scale street protests. The data are gathered in real time, do not necessitate advance preparation by researchers, and do not depend on either active cooperation or self-reporting on the part of protesters. As a result, for reasons explained in part 1 of this article, this type of data has a number of advantages over alternative sources for conducting quantitative analysis of participation in mass mobilizations.

No less important are the reasons why researchers should want to carry out empirical research of the kind that the new data make possible. The social justice protest in Israel in summer 2011 was part of a rare but recurrent phenomenon, the emergence (usually unanticipated) of mass mobilizations that spread within and across countries. If, as many scholars believe, the recent waves were facilitated by new modes and technologies of communication, future national and/or international campaigns featuring massive street protests are more likely than in the past. This adds to the importance of expanding the toolbox for conducting quantitative research on both the *composition* and *dynamics* of mobilization. As a result of sampling difficulties, data quality issues, and cost constraints, established approaches have serious limitations in investigating these two facets of large protest campaigns. While the location data used in this article for illustrative purposes have their own limitations (discussed below), they made it possible for us to identify the types of communities that sent protesters to the campaign and when, where, and with whom they participated.

There are also other reasons why research of this kind matters. In their seminal work *Dynamics of Contention*, McAdam, Tarrow, and Tilly (2001) set an agenda that called for focusing on *episodes of contention* and exploring the mechanisms and processes that explain their development. While the concept of protest *campaign* was not developed in their volume (cf. Tilly and Tarrow 2007), the diverse empirical cases that McAdam et al. brought to life are nearly all linked episodes of contention. This underlines the importance of adding to existing capabilities for carrying out research informed by a campaign-oriented perspective (Kriesi 2009). Such research provides broader context to particular events and also accounts for details that are often overlooked when fully-fledged social movements are studied. Scholars have specifically argued that using campaigns as the unit of analysis is vital for understanding the evolution of mass mobilizations and their dissolution as well (Andrews and Biggs 2006; Staggenborg and Lecomte 2009). In the context of future waves or cycles of contention (Koopmans 2004; Tarrow 2011), the availability of time-stamped and geo-tagged behavioral data will enable major advances in our understanding of the diffusion of mass protests.

Another unresolved question in protest studies concerns the arc of contentious collective action over the last half-century. As Van Aelst and Walgrave (2001) put it, we know that acts of protest have become "normalized," but is this also true of the type of people who protest? By their nature, mass protest campaigns are an ideal test bed for investigating limits to the socioeconomic, cultural, and political inclusiveness of mass street protests. Previous scholarship has identified two different conditions for inclusive mass mobilization: *Negative coalitions* sustained by perceptions of a common enemy (Beissinger 2013; Dix 1984), and *emotional movements* triggered by a sense of shock and outrage that transcends entrenched social and political cleavages (Walgrave and Verhulst 2006). Our empirical application of the uses of location data to study Israel's 2011 protest campaign operationalized inclusivity by measuring the engagement of social sectors that were least likely to participate (and also least researchable using conventional methods). At the same time, the ability to evaluate time and space variations in mobilization exposed new or rarely studied inclusionary mechanisms including spatially accessible events, "magnet" event locations, and group segmentation of protesters at the event level.

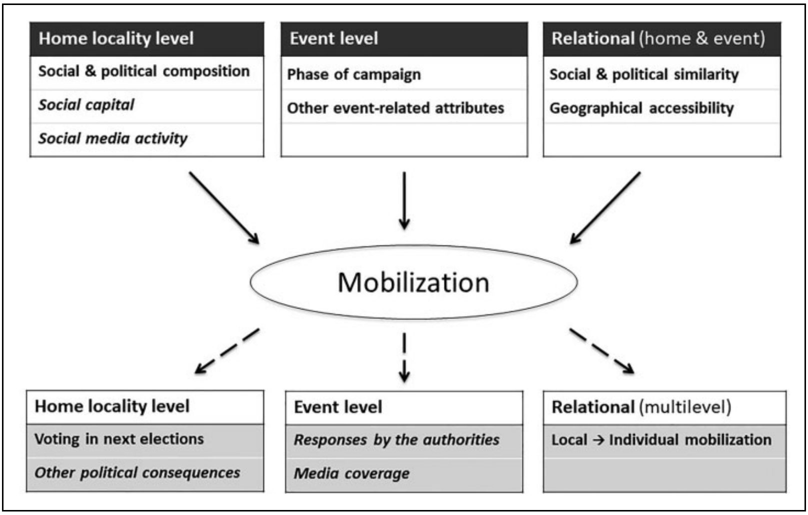

Figure 10. Testable effects using geographically aggregated protest participation data. Effects shown in italics are not assessed in this article.

The Uses of Mobile Location Data

This section proposes an analytical schema for thinking systematically about the range of research questions that can be tackled using the type of data set analyzed in this article. The schema has three main features:

- The relevant research questions are framed as testable effects.
- Two types of effects are distinguished, depending on whether protester mobilization serves as a dependent or independent variable.
- The aim is not only to map the specific types of analysis undertaken for this article but also to identify other potentially fruitful applications.

The schema is presented in Figure 10, where local mobilization (protest participation relative to the size of the local population) is positioned in the center. Mobilization is potentially explained by the independent variables shown above it and also serves as an independent variable for the purpose of explaining the outcomes shown below it. Consequently, the causal arrows coming from above (continuous lines) represent hypothesized *effects on protest*, while those pointing below (broken lines) signify *effects of protest*.

Both the upper and lower levels of Figure 10 are divided into three columns, representing three different levels of analysis. Our comments begin with effects *on* protest (the upper level) and move from left to right.

Analyses in the first column test how *attributes of home localities* affect local mobilization. In our research, the impact of the political orientation of the residents was especially noteworthy.

The second column refers to effects of *event characteristics* on the breadth and composition of mobilization. Findings reported earlier showed how the timing and framing of events influenced mobilization in the Israeli protest. Future research could examine such effects systematically by constructing indicators of the discursive framing of specific events by protest organizers and other relevant actors—elites, opponents, the mass media, and social media networks.

It is reasonable to expect that different phases and sites of protest attract different types of protesters (cf. Soule 2004). The third column, described as relational, addresses the questions *where, when*, *and with whom* mobilization occurred by connecting the places where protesters live to the times and places at which they participate and by attending to both the social and spatial distance between home and event localities. Previously reported findings for Israel have emphasized the role of spatial and temporal segmentation in facilitating the mobilization of diverse publics.

The *effects of protest* shown in the lower level of Figure 10 follow the same logic. The first column refers to possible downstream effects of rates of local mobilization on subsequent political outcomes. Vote shifts in the wake of the Israeli protest appear to support this expectation. The second column asks whether the number and composition of protest participants at specific events influences the response of other salient actors. The third column draws attention to the multilevel relationship between *aggregate* mobilization (measured behaviorally at the level of home localities) and the propensity of *individuals* to join the protest (measured by a survey that samples the same localities). Our research indicates that in Israel, the likelihood of participation by otherwise similar individuals was indeed boosted by the level of mobilization in their community—providing they were not politically estranged from the protest core.

The five effects shown in italics in Figure 10 were not explored in our analyses. The box on the upper left-hand side includes two suggestions. One is to consider the role of local social capital in the propensity of residents to participate in mass demonstrations. The link between social capital at the neighborhood level and engagement in politics and collective action has long been discussed (Krishna 2002; Putnam 2001; Verba, Schlozman, and Brady 1995), with some scholars specifically pointing to its role in the emergence of social movements (e.g., Nicholls 2009; Polese 2009). For the purpose of testing the effects of social capital on protest propensity, geographically aggregated data on protest are arguably preferable to the more commonly used individual-level measures of participation, assuming that social capital is theorized as a property of communities rather than individuals.

A further issue that lies beyond the scope of this article is the relationship between online and off-line mobilization in mass protests. Studies of two notable campaigns of 2013—in Turkey (Barberá et al. 2015) and Brazil (Bastos, da Cunha Recuero, and da Silva Zago 2014)—revealed geographical separation between online supporters and off-line demonstrators. These findings imply a *complementary* division of labor between the two spheres of participation in contemporary large-scale protest campaigns. Studies in other settings have probed the *sequential* relationship between local online and off-line activity. Alongside evidence that social media activity in a particular area is followed by street protest (Steinert-Threlkeld 2017), there are studies suggesting either that the two trend in parallel (Jungherr and Jürgens 2014) or that local mobilization spurs subsequent online activity (Porto and Brant 2015:197). Further advances in this area would be greatly facilitated by utilizing mobile location data, which provide far more precise spatial and temporal coordinates of participation than the sources on which previous research has relied. Specifically, analysis of online–off-line interactions would benefit from linking geotagged social media activity with location data for active participants in street demonstrations.

The box on the lower left side of Figure 10 notes that the potentially testable downstream effects of protest on political action are not confined to election outcomes. If locality-level data are available, it would also be possible to analyze other political legacies of a mass mobilization campaign. For example, in what types of communities was mobilization followed by a higher probability of participation in later outbreaks of protest, the strengthening of local civil society organizations and the emergence of new ones, or changes in local public policy?

Moving to the event level (center column), many additional event-level influences on mobilization that have been identified by social movement researchers could be profitably investigated. They include the sponsoring organizations, the tactics deployed, and the presence of police and counter-demonstrators. Several downstream effects of mobilization, capturing the dynamics of protest events, are suggested at the lower level. Are responses by the authorities, such as police repression, influenced by the number and identity of the protesters? How much are the extent and tone of mass media coverage affected by who protests?

The framework in Figure 10 is not intended to be exhaustive. It is our hope that it will stimulate thinking about other potential uses for the kind of data analyzed in this article. Specifically, Figure 10 does not address the potential for exploiting locality-based data on mobilization by using concepts and methods from disciplines outside sociology or political science. In particular, geography and network studies are equipped with concepts, statistical tools, and methods of visualization capable of generating insights into the spatial distribution of protesters, flows between home and event locations, and the spatiotemporal diffusion of protest.

Limitations of the Data

The many merits of mobile location data for researching mass street protests do not mean they are without flaws or that we can expect other methodologies to become redundant.[27] For one thing, location data are unsuitable for monitoring forms of collective action that are often deployed in conjunction with demonstrations such as online political action, protest camps, small demonstrations, and strikes. Still, we have argued that even without being able to incorporate data on the broader repertoire of contention, there are good reasons to study large-scale demonstrations. Their visibility, contribution to total protest participation, and potential impact on discourse, policy, and power are unlikely to be matched by other forms of collective contentious action. For this reason, street demonstrations remain the flagship form of action by contemporary mass movements.[28]

In the context of street protests, mobile location data alone cannot offer a comprehensive portrait of who joins collective acts of protest and why. True, the behavioral nature of the data automatically generated by telecommunications and smartphones is a major strength. Nevertheless, the invisibility of subjectivity means that no information is available on the motivations, perceptions, intentions, and emotions of the actors who participate or refrain from participating in large demonstrations. These absences limit the range of research questions that can be addressed and leave out many of the social, psychological, and political mechanisms which scholars have suggested can explain whether and why individual protesters mobilize. Additional research methods, both quantitative and qualitative, will therefore continue to play an essential role.

Data availability could also be an issue. A scenario in which the authorities are suspected by protesters of using location data to identify and punish them may cause participants to leave their mobile phones at home or turned off during demonstrations. Other limitations depend on the specifics of how location data are collected and disseminated by individual cellular networks and data vendors and the restrictions imposed by regulators. For example, when either networks or intermediaries perform in-house analysis for clients, they work with the original locational data, typically merged with other individual-level data. In contrast, researchers wishing to carry out their own analyses may be limited to data extracts that are less rich and less precise.

Another caveat is that while mobile location data improve our ability to estimate the size of protests, they have serious limitations in this respect. It has been shown that several types of location data are capable of supplying quite accurate estimates of crowd size (Botta, Moat, and Preis 2015) but under restrictive conditions that did not apply in our case and would also not apply in others. The event of interest would have to take place in a clearly bounded area, individual-level data would be continuously available in order to deal with the problem of turnover (people arriving and departing over the course of an event), and the target crowd would not include bystanders. As pointed out when introducing our data set, distinguishing bystanders from protesters is arguably the most serious challenge to accurately identifying participants in most mass protest events.[29] In the absence of explicit information on intentionality at the individual level, researchers need to focus on improving the precision with which protest spaces are defined and the crowds within them are counted and on developing enhanced methods of indirectly assessing the likely presence of bystanders. Uncertainty regarding the dividing line between protesters and nonprotesters introduces an unquantifiable element of uncertainty into analyses like those conducted for this article.

The geographical aggregation of our data is convenient in some respects but also imposes constraints. The range of covariates which it is feasible to analyze excludes variables (e.g., gender) that cannot be meaningfully aggregated by geographical units. Even with suitable attributes, such as the social and political variables used in our examples, interpretation of relationships between the compositional features of home localities and their rate of mobilization runs the risk of erroneous causal inference. The so-called ecological fallacy arises because associations between aggregated attributes of the people living in a given geographical unit and their presumed consequences may be quite different to those found when the same associations are tested between individuals (Glynn and Wakefield 2010). That said, there are a number of ways in which the risk of ecological fallacy can be assessed and potentially reduced.[30]

Finally, the more routine (but no less vital) data quality issues touched on when we introduced our data set in part 2 need to be considered. Are phone owners in general, and the specific sample of owners whose devices are tracked by a given data vendor, representative of the population of potential protesters? If not, are the gaps successfully closed when estimates are weighted and extrapolated to population level? How reliable is the assignment of devices to home localities?

The standard scientific procedure for evaluating the seriousness of such issues is external validation, something which Calabrese, Ferrari, and Blondel (2014) have urged researchers using location data to undertake. In a rare reported instance in which validation was carried out, the researchers conducted a study that utilized location data supplied by our vendor. Their study of commuting patterns throughout Israel yielded results described as "remarkably consistent" with census data (Razin and Charney 2015:1142). Unfortunately, for reasons already mentioned in the case of protest participation, no ground truth data are available. Lacking the option of external validation against a gold standard, we have offered sensitivity tests as an alternative way to assess the robustness of substantive findings (Online Appendix 2, which can be found at http://smr.sagepub.com/supplemental/).

The potential limitations to which we have drawn attention should certainly not deter protest researchers from utilizing mobile location data for studying large protests. The drawbacks of any method of data collection can only be judged in relation to its benefits and the availability of alternatives. In our judgment, the advantages of the data introduced and analyzed here far outweigh their limitations. Moreover, the field of locational big data is rapidly expanding, and with this expansion have come more precise and continuous location recording and enhanced integration of complementary information, often at the individual level. That said, the increasing accuracy, richness, and specificity of location data add to the ethical concerns they raise, which is the final topic we address.

Privacy and Ethical Issues

The most obvious ethical challenge for researchers using mobile location data is obtaining informed consent from phone owners whose movements are monitored.[31] The fine print of contracts signed by the subscribers of a cellular network often includes agreeing to the network operator storing and selling their location data. Similarly, smartphone apps are required (at least on first use) to request permission to activate location services. It appears however that most users are unaware of the extent to which their location is tracked, and the resulting information is resold for profit and made available to the authorities. Even those aware of the intrusive nature and potentially pernicious uses of location data may be intimidated by the difficulty of understanding and enabling device and application settings that limit or prohibit tracking by location services. They may also be reluctant to forfeit the usability and usefulness that are lost when apps are denied access to these services.

To protect the privacy of individual phone owners, governments impose varying levels and types of regulation. Vendors and analysts of data generated by the standard operating procedures of cellular networks may be obligated to anonymize the identity of devices in their own database and to supply clients with data sets containing aggregated data. However, these precautions in themselves do not necessarily prevent individuals from being identified. That depends in large part on the strictness of the rules applied such as how frequently the internal ID numbers of devices have to be anonymized and how much the data supplied to end users are blurred.

The preparation and sale of location information gathered by smartphone apps is more challenging to control and is also less constrained by regulation (at least in the United States) than data derived from mobile network operators.[32] Reasons for this include the relative newness and evolving nature of the location services data industry and the implicit expectations of policy makers that corporations like Google and Facebook will self-regulate or smartphone owners will grasp and responsibly exercise their ability to protect their own privacy. Ancillary data supplying rich complementary information on individuals add an even more serious privacy risk. This type of information is offered by major cell network operators, not only shady data aggregators.[33]

The difficulty of respecting the privacy of mobile phone owners, especially at a time when the invasiveness of data gathering is increasing along with the richness and precision of available data, poses the risk that academic researchers may find themselves barred altogether by funding agencies or their own institutions from utilizing mobile location data for research. Yet, given the nature of automatically collected data, the many hands through which it may pass, and the constantly expanding uses for which such data are being exploited, it could be argued that for this type of data, user consent could never be more than dimly informed. Consequently, we believe that efforts to protect the unknowing subjects of research based on big location data should focus on building privacy safeguards into the data generation process, maximizing deterrence of privacy violations by data vendors, and placing barriers between researchers and individual-level data.

Where governments fail to meet these goals, academic institutions might adopt their own regulations aimed at creating an acceptable balance between researcher access and ethical considerations. Two types of privacy-enhancing barriers would help ease the trade-off. One is to follow the lead of official statistical bureaus that create a sterile space between researchers and sensitive micro-data sets by providing closely regulated access in secure research spaces. In addition, researchers could be offered the option of unmediated access to "safe" data sets like the one we acquired, consisting of count data that have been preaggregated by geographical home locations (cf. Sanches et al. 2014). The conviction that this type of data set can substantially advance knowledge without risking the privacy and wellbeing of unwitting phone owners is an important reason why this article has gone into such detail concerning the preparation and processing of locality-aggregated data sets and the specific ways in which they can advance social research on street protests.

Finally, we recommend that researchers of mass protest prepare themselves to deploy a multipronged approach to location data when future mass protest events take place. For example, the type of thin-but-wide data on which this article has focused could be productively combined with supplementary data based on voluntary location tracking by phone owners who are either found at or directed to the site of a demonstration. Using a special-purpose smartphone app, these informants could be asked to activate their GPS and periodically provide real-time feedback on their intentions, perceptions, and subjective experiences (cf. Palmer et al. 2013).[34] The resulting deep-but-narrow data would make it possible for researchers to fill important gaps and would add great value to the type of research reported in this article.

In closing, we wish to succinctly reiterate the two different goals of this article. First, to alert readers to the truly new and exciting contributions that mobile location data can and will make, not only to the study of large-scale street protests but also to resolving debates and addressing lacunae that have broad significance for the study of contentious collective action. Our second aim has been to encourage the development of skills and procedures that will facilitate undertaking and enhancing research of this kind in the future. We hope that our work will contribute to awareness of both old and new research questions that location data can address; technical desiderata, pitfalls and trade-offs, and their implications; the development of transparent procedures for data cleaning and processing that will eventually become standardized; and familiarity with data analysis tools that we have found helpful. Continued pooling of expertise and experience in the future will do much to realize the potential of this new field of inquiry. Researchers wishing to utilize our own data set will be able to do so under conditions of phased availability.[35]

Authors' Note

We received valuable feedback at a workshop we convened at Tel Aviv University and in seminars at Sciences Po (Sociology), Berkeley (Political Science), and NYU (Sociology).

Acknowledgments

We are grateful to Israel's Ministry of Science and Technology for a three-year grant to Michael Shalev and Yossi Shavit. This research would not have been possible without Yossi's contributions, both substantive and intellectual, or Adi Ben-Nun's help in obtaining the data and sharing his expertise. We also thank Reut Shafrir for valuable research assistance and the many scholars who have generously provided advice and constructive criticism. Hanspeter Kriesi, Michael Biggs, Mario Diani, Amit Birenboim, and the anonymous reviewers for *SMR* made especially important contributions.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was funded by Ministry of Science and Technology, Israel (Grant ID: 3-3473).

ORCID iD

Michael Shalev https://orcid.org/0000-0003-3308-6368

Supplemental Material

The supplemental material for this article is available online (https://journals.sagepub.com/doi/suppl/10.1177/0049124120914926).

Notes

1. A protest campaign is a "thematically, socially, and temporally interconnected" series of contentious collective actions (Della Porta and Rucht 2002:3). Campaigns have typically been conceived as part of the repertoire of fully-fledged

social movements, but as recent instances of mass protest make clear, they may also be initiated by activists with connections to a variety of movements or none at all (Flesher Fominaya 2014).
2. Examples include Beissinger, Jamal, and Mazur (2015), Rüdig and Karyotis (2014), and Yoruk and Yuksel (2014).
3. A recent comprehensive guide to methods and applications for using twitter as social science data by Steinert-Threlkeld (2018) includes helpful information on the quality of twitter-derived location data and its potential uses. Our reading of the literature (e.g., Cesare, Grant, and Nsoesie 2017; Graham, Hale, and Gaffney 2014; Malik et al. 2015; Sloan and Morgan 2015) is that twitter location data are generally less robust than mobile location data. Only a small proportion of tweets include metadata or content that can be used to infer their location with comparable or better accuracy, samples of tweets can be biased if the service is not universally used (a significant limitation in some countries, including Israel in 2011), and the background information available on tweeters is limited (although new methods are being developed to improve this capability). That said, a recent study of participation in the Women's Marches held throughout the United States following the 2016 Presidential election finds that tweet-based estimates of the *order of magnitude* of protest events were highly correlated with estimates derived from mobile location data (Sobolev et al. 2019).
4. The differences include location accuracy, frequency of data collection, vulnerability to network overload, whether and under what conditions the data are available to end users, and what linked ancillary data (if any) are provided. Since the details vary between networks (depending on their hardware and software), as well as between countries and vendors, they can only be established on a case-by-case basis.
5. The sole published study of protest activity based on location data of which we are aware, by Traag, Quax, and Sloot (2017), is quite different from the present research. The researchers obtained an entire year of individual call detail records from an unidentified African country, which they matched to 39 small standalone protest events (mean = 340 participants).
6. The data vendor was TrendIt, an Israeli technology company active between 2008 and 2016. The mobile network carrier generating the data was Cellcom, which in the year of the protest accounted for 35.7 percent of all Israeli subscribers (Cohen 2013), and was considered to have especially strong coverage among two large minorities (Arab citizens and ultra-orthodox Jews) which have proven difficult to investigate with surveys (Harel Kfir 2011; Kristal 2011). We purchased data from TrendIt after being informed that the company's data gathering and distribution practices were governed by regulations of Israel's Ministry

of Communications, which protect the identity of network subscribers and forbid the sale of data which could be used to identify individuals.

7. Central Bureau of Statistics. Retrieved March 30, 2017 (http://www.cbs.gov.il/mifkad/mifkad_2008/hagdarot_e.pdf).
8. The original data extraction included seven events that we later disqualified. Two of these were not integral to the protest campaign. One small event (with a reported size of 2,000) was excluded after discovering that the vendor had mistakenly used a target area provided for a different city. The remaining four events were excluded because they yielded estimates of fewer than 1,000 participants. Note that three other events with fewer than 1,000 participants were retained because they occurred at locations that also hosted additional events, enabling their campaign-wide participations to easily pass our threshold.
9. For reasons explained in Online Appendix 1 (which can be found at http://smr.sagepub.com/supplemental/), the procedure we adopted is more relevant to bystanders from outside an event locality than to local residents.
10. Note that the sample does not include 53 home localities with missing data on covariates. Population data (and all other data utilized in this research except for election results) are from Israel's 2008 census. They cover all residents of Israel, including Jewish settlers in the occupied territories. We excluded Palestinian residents of East Jerusalem because they are not Israeli citizens and do not participate in Israeli politics. In any event, they could not have been meaningfully included in the analysis since the Central Bureau of Statistics treats the whole of Arab East Jerusalem as a single statistical area.
11. The word participations is more accurate than participants, since the same individuals may have participated in more than one event.
12. Extreme values of peak participation (in excess of 80 percent of the working-age population), found in only 28 home localities, were tapered to a top-coded upper limit (100 percent) while preserving their rank order.
13. According to surveys conducted during the protest, a vast majority of the public were sympathetic to the protest. For example, the *Peace Index* surveys carried out in July and August 2011 (available from www.peaceindex.org/defaultEng.aspx) indicate that it enjoyed the support of (respectively) 88 percent and 78 percent of the general public.
14. Henceforth "participation rate" will be used as shorthand for "peak participation rate", defined earlier as the highest hourly number of presumed protesters recorded from a home locality throughout the campaign, relative to its working-age population.
15. Estimates of the ultra-orthodox population are from Malach, Choshen, and Cahaner (2016). Figures for the other two sectors are drawn from the *Statistical Abstract of Israel 2015* (tables 2.1 and 2.16). Palestinians resident

in East Jerusalem (who are predominantly not citizens) are not included in the analysis, and settlers do not include the large number of Jewish residents of Jerusalem living in parts of the city that were conquered in 1967. Note that due to above-average birthrates, especially among the ultra-orthodox, the minority sectors' share of potential protesters is significantly smaller than their share of the population.

16. For an overview of social and political cleavages in Israel, see Sasley and Waller (2016).
17. Detailed information on the per-sector distribution of participation rates across active home localities is provided in Online Appendix 4 (which can be found at http://smr.sagepub.com/supplemental/). All distributions follow a similar log-normal form.
18. The first two of the cited studies relied on self-reports by survey respondents of their distance from the nearest protest event, whereas the third (by Traag et al.) utilized mobile location data.
19. Arab home localities are included in this analysis, but many are missing data on the distribution of votes. We dealt with this problem by assigning all Arab localities a token left-center vote share of 1 percent, randomly jittered.
20. Multiple fractional polynomial modeling is a flexible method that iteratively identifies the polynomial function that best fits the association between each covariate and the dependent variable. By introducing power transformations of the predictor variables, this method can successfully model complex and non-linear relationships and provide estimates that closely reflect the underlying data. (for more details, see Royston and Sauerbrei 2008; Sauerbrei and Royston 2016).
21. However, due to diminishing sample sizes, at distances exceeding around 20 kilometers the magnitude of the net minority–majority gap becomes statistically uncertain.
22. For example, localities that are situated 40 kilometers away from the most accessible protest are 49 percent (majority) and 61 percent (minority) less likely to be active than their counterparts at a distance of 30 kilometers.
23. Participation rates are logged because their distribution is skewed to the left and has a long right tail.
24. The single exception is that in the second cycle settler participation hardly rose at the peak.
25. See Online Appendix 5 (which can be found at http://smr.sagepub.com/supplemental/) for further information on the 2012 "Democracy Index" survey utilized for this analysis. Since the survey geocoded municipalities rather than home localities (i.e., census tracts), we aggregated estimates of local mobilization to the municipal level before merging them with survey data for 659 geocoded respondents residing in 118 municipalities.

26. A different threat to the causal status of the effect of local mobilization, which we do not attempt to address here, is that individuals may self-select into localities on grounds that include or correlate with local participation norms.
27. For general discussions of the limitations of big data, see Boyd and Crawford (2012) and Tufekci (2014). Mobilizing Ideas (2015) published an online symposium on big data in social movement research.
28. Occupy Wall Street, which focused almost entirely on encampments and online activity, is a notable exception.
29. The Women's March, referenced at the outset of this article, is an example of a mass event held in a clearly demarcated area. Nevertheless, some of those present at such an event could still be bystanders of a kind, either curious but not committed or agents provocateurs. It bears emphasis that the problem of bystanders is inherent in all indirect methods of identifying and counting protesters, whether observational (geo-located data or traditional "crowd science" methods) or inferential (based on the content of text and images posted on social media).
30. In work in progress using the Israeli data set analyzed in this article, we adopt three approaches. First, as an alternative to attempting to isolate the net effects of specific locality attributes via some form of multivariate analysis, we cluster multiple attributes of localities into distinct configurations or "types" using latent class analysis (Magidson and Vermunt 2004). Second, we are experimenting with an approach proposed by Rosen et al. (2001) for placing logical and statistically reasonable bounds on ecological associations. Third, in order to improve ecological inferences, we intend to make use of multilevel data sets that combine individual and aggregate data, like the one used in this article to investigate contextual effects on individuals (Jackson, Best, and Richardson 2006).
31. The summary that follows is based on research articles cited in part 1 of this article (mainly in the subsection Sources of Mobile Location Data), other specialized academic papers (Jiow 2016; Zurbarán et al. 2014), and investigations by news organizations and privacy watchdogs (e.g., Valentino-Devries et al. 2018).
32. Effective May 2018, the European Union introduced the General Data Protection Regulation, which among other things allows heavy penalties to be imposed on digital information companies that violate user rights to location privacy.
33. In May 2017, the data division of one of the largest cellular networks in the UK proposed selling one of the authors a data set for analyzing participation in large protest events in the UK since 2013, that would include personal attributes (age and gender) culled from subscriber records, and weblogs and app usage data scraped from their devices. The proposal incongruously added that "We are required by data privacy guidelines to aggregate to units of 10."

34. This type of interactive location tracking using a known sample would also be effective in testing the external validity of indirect methods of identifying bystanders and improving their accuracy.
35. The working data set can be obtained from the authors for replication purposes. Two years after publication of this article, the data set will be made available for any purpose. Kindly contact the corresponding author.

References

Ahas, Rein. 2011. "Mobile Positioning." Pp. 183-99 in *Mobile Methods*, edited by M. Büscher, J. Urry, and K. Witchger. London, England: Routledge.

AlSayyad, Nezar and Muna Guvenc. 2015. "Virtual Uprisings: On the Interaction of New Social Media, Traditional Media Coverage and Urban Space during the 'Arab Spring'." *Urban Studies* 52:2018-34.

Ancelovici, Marcos, Pascale Dufour, and Héloïse Nez. 2016. *Street Politics in the Age of Austerity: From the Indignados to Occupy*. Amsterdam, the Netherlands: Amsterdam University Press.

Andretta, Massimiliano and Donatella Della Porta. 2014. "Surveying Protestors: Why and How." Pp. 397-417 in *Methodological Practices in Social Movement Research,* edited by D. Della Porta. Oxford, England: Oxford University Press.

Andrews, Kenneth T. and Michael Biggs. 2006. "The Dynamics of Protest Diffusion: Movement Organizations, Social Networks, and News Media in the 1960 Sit-ins." *American Sociological Review* 71:752-77.

Anduiza, Eva, Camilo Cristancho, and José Manuel Sabucedo. 2014. "Mobilization through Online Social Networks: The Political Protest of the Indignados in Spain." *Information, Communication & Society* 17:750-64.

Anduiza, Eva, Irene Martin, and Araceli Mateos. 2013. "The Consequences of the 15-M Protests on Voting Choice in the 2011 Spanish General Elections." EPSA 2013 Annual General Conference Paper 351. Retrieved May 4, 2014 (http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2224814).

Aziz, Muhammad Waqar, Farhan Naeem, Muhammad Hamad Alizai, and Khan Bahadar Khan. 2018. "Automated Solutions for Crowd Size Estimation." *Social Science Computer Review* 36:610-31.

Barberá, Pablo, Ning Wang, Richard Bonneau, John T. Jost, Jonathan Nagler, Joshua Tucker, and Sandra González-Bailón. 2015. "The Critical Periphery in the Growth of Social Protests." *PLoS One* 10:e0143611.

Bastos, Marco Toledo, Raquel da Cunha Recuero, and Gabriela da Silva Zago. 2014. "Taking Tweets to the Streets: A Spatial Analysis of the Vinegar Protests in Brazil." *First Monday* 19. doi: 10.5210/fm.v19i3.5227.

Beissinger, Mark R. 2013. “The Semblance of Democratic Revolution: Coalitions in Ukraine’s Orange Revolution.” *American Political Science Review* 107: 574-92.

Beissinger, Mark R., Amaney A. Jamal, and Kevin Mazur. 2015. “Explaining Divergent Revolutionary Coalitions: Regime Strategies and the Structuring of Participation in the Tunisian and Egyptian Revolutions.” *Comparative Politics* 48:1-24.

Bennett, W. Lance and Alexandra Segerberg. 2013. *The Logic of Connective Action: Digital Media and the Personalization of Contentious Politics*. Cambridge, England: Cambridge University Press.

Beyerlein, Kraig, Peter Barwis, Bryant Crubaugh, and Cole Carnesecca. 2016. “A New Picture of Protest: The National Study of Protest Events.” *Sociological Methods & Research* 1-46. doi: 10.1177/0049124116661574.

Biggs, Michael. 2016. “Size Matters: Quantifying Protest by Counting Participants.” *Sociological Methods & Research* 1-33. doi: 10.1177/0049124116629166.

Botta, Federico, Helen Susannah Moat, and Tobias Preis. 2015. “Quantifying Crowd Size with Mobile Phone and Twitter Data.” *Royal Society Open Science* 2. Retrieved May 24, 2019 (https://doi.org/10.1098/rsos.150162).

Boyd, Danah and Kate Crawford. 2012. “Critical Questions for Big Data.” *Information, Communication & Society* 15:662-79.

Braun, Robert and Ruud Koopmans. 2010. “The Diffusion of Ethnic Violence in Germany: The Role of Social Similarity.” *European Sociological Review* 26: 111-23.

Calabrese, Francesco, Laura Ferrari, and Vincent Blondel. 2014. “Urban Sensing Using Mobile Phone Network Data: A Survey of Research.” *ACM Computing Surveys* 47:1-20.

Cesare, Nina, Christan Grant, and Elaine O. Nsoesie. 2017. “Detection of User Demographics on Social Media: A Review of Methods and Recommendations for Best Practices.” arXiv preprint 1412.2247.

Cohen, Erez. 2013. “Intervention or Involvement: Regulation Processes in the Cellular Communication Industry in Israel (Hebrew).” *Misgarot Media* 11:79-117.

Converse, Philip E. and Roy Pierce. 1986. “Revolt and Reaction: The Upheaval of May 1968 and the June Elections.” Pp. 417-43 in *Political Representation in France*, edited by Philip Converse and Roy Pierce. Cambridge, MA: Harvard University Press.

Dantcheva, Antitza, Petros Elia, and Arun Ross. 2016. “What Else Does Your Biometric Data Reveal? A Survey on Soft Biometrics.” *IEEE Transactions on Information Forensics and Security* 11:441-67.

Della Porta, Donatella. 2015. *Social Movements in Times of Austerity: Bringing Capitalism Back into Protest Analysis*. Cambridge, England: Polity.

Della Porta, Donatella, and Dieter Rucht. 2002. "The Dynamics of Environmental Campaigns." *Mobilization* 7:1-14.

Della Porta, Donatella. 2017. "Riding the Wave: Protest Cascades, and What We Can Learn from Them." Pp. 9-30 in *Global Diffusion of Protest: Riding the Protest Wave in the Neoliberal Crisis*, edited by D. Della Porta. Amsterdam: Amsterdam University Press.

Dix, Robert H. 1984. "Why Revolutions Succeed & Fail." *Polity* 16:423-46.

Flesher Fominaya, Cristina. 2014. *Social Movements and Globalization: How Protests, Occupations and Uprisings Are Changing the World*. London, England: Palgrave Macmillan.

Fox, Ryan Squire, and Victoria Gu. 2017. "Inauguration Attendees Make Significantly Less Money than Women's March Attendees." January 27. Retrieved December 12, 2018 (https://blog.safegraph.com/inauguration-attendees-make-significantly-less-money-than-womens-march-attendees-7cb8b056556a).

Gelman, Andrew, Boris Shor, Joseph Bafumi, and David Park. 2007. "Rich State, Poor State, Red State, Blue State: What's the Matter with Connecticut?" *Quarterly Journal of Political Science* 2:345-67.

Gerbaudo, Paolo. 2017. *The Mask and the Flag: Populism, Citizenism and Global Protest*. Oxford, England: Oxford University Press.

Gillion, Daniel Q. and Sarah A. Soule. 2018. "The Impact of Protest on Elections in the United States." *Social Science Quarterly* 99:1649-664.

Glynn, Adam N. and Jon Wakefield. 2010. "Ecological Inference in the Social Sciences." *Statistical Methodology* 7:307-22.

Goldstone, Jack A. 2011. "Cross-class Coalitions and the Making of the Arab Revolts of 2011." *Swiss Political Science Review* 17:457-62.

Graham, Mark, Scott A. Hale, and Devin Gaffney. 2014. "Where in the World Are You? Geolocation and Language Identification in Twitter." *The Professional Geographer* 66:568-78.

Haber, Carmit, Ella Heller, and Tamar Hermann. 2011. *Solidarity in the Protests of Summer 2011 (Hebrew)*. Jerusalem, Israel: Israel Democracy Institute.

Harel Kfir, Daphna. 2011. "For the Retailers' Information: The Consumers in the Arab Sector Are Undergoing Israelization (Hebrew)." Globes 16.6.2011. Retrieved May 31, 2014 (http://www.globes.co.il/news/article.aspx?did=1000654547).

Henke, Lucy L. 2016. "Estimating Crowd Size: A Multidisciplinary Review and Framework for Analysis." *Business Studies Journal* 8:27-38.

Huber, George P. and Danial J. Power. 1985. "Retrospective Reports of Strategic-level Managers: Guidelines for Increasing Their Accuracy." *Strategic Management Journal* 6:171-80.

Huckfeldt, R. Robert. 1979. “Political Participation and the Neighborhood Social Context.” *American Journal of Political Science* 23:579-92.

Hutter, Swen. 2014. “Protest Event Analysis and Its Offspring.” Pp. 335-67 in *Methodological Practices in Social Movement Research*, edited by D. Della Porta. Oxford, England: Oxford University Press.

Jackson, Christopher, Nicky Best, and Sylvia Richardson. 2006. “Improving Ecological Inference Using Individual-level Data.” *Statistics in Medicine* 25:2136-159.

Jiow, Hee Jhee. 2016. “Mobile Location Based Services: Implications on Privacy.” Pp. 765-78 in *Geospatial Research: Concepts, Methodologies, Tools, and Applications,* edited by Information Resources Management Association. Hershey, PA: Information Science Reference.

Joo, Jungseock and Zachary C. Steinert-Threlkeld. 2018. “Image as Data: Automated Visual Content Analysis for Political Science.” arXiv preprint arXiv:1810.01544.

Jungherr, Andreas and Pascal Jürgens. 2014. “Stuttgart’s Black Thursday on Twitter: Mapping Political Protests with Social Media Data.” Pp. 154-96 in *Analysing Social Media Data and Web Networks*, edited by R. Gibson, M. Cantijoch, and S. Ward. Houndmills, England: Palgrave Macmillan.

Karp, Jeffrey A. and David Brockington. 2005. “Social Desirability and Response Validity: A Comparative Analysis of Overreporting Voter Turnout in Five Countries.” *Journal of Politics* 67:825-40.

Klandermans, Bert. 1997. *The Social Psychology of Protest*. Oxford, England: Blackwell.

Koopmans, Ruud. 2004. “Protest in Time and Space: The Evolution of Waves of Contention.” Pp. 19-46 in *The Blackwell Companion to Social Movements*, edited by D. A. Snow, S. A. Soule, and H. Kriesi. Oxford, England: Blackwell.

Kriesi, Hanspeter. 2009. “Charles Tilly: Contentious Performances, Campaigns and Social Movements.” *Swiss Political Science Review* 15:341-49.

Krishna, Anirudh. 2002. “Enhancing Political Participation in Democracies.” *Comparative Political Studies* 35:437-60.

Kristal, Merav. 2011. “Cellcom Is the Leading Cellular Provider in the Ultra-Orthodox Sector (Hebrew).” Ynet 30.1.2011. Retrieved May 15, 2014 (http://www.ynet.co.il/articles/0,7340,L-4021255,00.html).

Magidson, Jay and Jeroen K. Vermunt. 2004. “Latent Class Models.” Pp. 175-98 in *The Sage Handbook of Quantitative Methodology for the Social Sciences*, edited by D. Kaplan. Thousand Oaks, CA: Sage.

Malach, Gilad, Maya Choshen, and Lee Cahaner. 2016. *Statistical Report on Ultra-Orthodox Society in Israel: 2016 Abstract*. Jerusalem, Israel: Jerusalem Institute for Israel Studies and Israel Democracy Institute.

Malik, Momin, Hemank Lamba, Constantine Nakos, and Jürgen Pfeffer. 2015. “Population Bias in Geotagged Tweets.” Pp. 18-27 in *ICWSM Workshop on*

Standards and Practices in Large-scale Social Media Research. Retrieved October 30, 2018 (https://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10662).
Mann, Justin. 2017. "The 2017 Women's March on Washington Had 3× the Attendance at Trump's Inauguration." *Fysical Blog*, December 29. Retrieved December 16, 2018 (https://medium.com/fysicalblog/the-2017-womens-march-in-washington-had-3x-the-attendance-as-trump-s-inauguration-267fc7f6968c).
Marpsat, Maryse and Nicolas Razafindratsima. 2010. "Survey Methods for Hard-to-reach Populations: Introduction to the Special Issue." *Methodological Innovations Online* 5:3-16.
McAdam, Doug and Sidney Tarrow. 2010. "Ballots and Barricades: On the Reciprocal Relationship between Elections and Social Movements." *Perspectives on Politics* 8:529-42.
McAdam, Doug, Sidney Tarrow, and Charles Tilly. 2001. *Dynamics of Contention*. Cambridge, England: Cambridge University Press.
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(3):478-99.
Mobile Marketing Association. 2015. "Demystifying Location Data Accuracy." Retrieved December 18, 2018 (http://www.mmaglobal.com/files/documents/location-data-accuracy-v3.pdf).
Mobilizing Ideas. 2015. "Social Movement Data (2-Part Symposium on Big Data)." Retrieved August 15, 2017 (https://mobilizingideas.wordpress.com/category/essay-dialogues/social-movement-data).
Myers, Daniel. 2010. "Violent Protest and Heterogeneous Diffusion Processes: The Spread of US Racial Rioting from 1964 to 1971." *Mobilization* 15:289-321.
Nicholls, Walter. 2009. "Place, Networks, Space: Theorising the Geographies of Social Movements." *Transactions of the Institute of British Geographers* 34:78-93.
Onuch, Olga. 2014. *Mapping Mass Mobilization: Understanding Revolutionary Moments in Argentina (2001) and Ukraine (2004)*. New York: Palgrave Macmillan.
Palmer, John R. B, Thomas J. Espenshade, Frederic Bartumeus, Chang Y. Chung, Necati Ercan Ozgencil, and Kathleen Li. 2013. "New Approaches to Human Mobility: Using Mobile Phones for Demographic Research." *Demography* 50:1105-128.
Perugorría, Ignacia, Michael Shalev, and Benjamín Tejerina. 2016. "The Spanish Indignados and Israel's Social Justice Movement: The Role of Political Cleavages in Two Large-Scale Protests." Pp. 97-124 in *Street Politics in the Age of Austerity:*

From the Indignados to Occupy, edited by M. Ancelovici, P. Dufour, and H. Nez. Amsterdam, the Netherlands: Amsterdam University Press.

Pierce, Roy and Philip E. Converse. 1990. "Attitudinal Sources of Protest Behavior in France: Differences between before and after Measurement." *Public Opinion Quarterly* 54:295-316.

Polese, Abel. 2009. "Ukraine 2004: Informal Networks, Transformation of Social Capital and Coloured Revolutions." *Journal of Communist Studies and Transition Politics* 25:255-77.

Porto, Mauro P. and João Brant. 2015. "Social Media and the 2013 Protests in Brazil: The Contradictory Nature of Political Mobilization in the Digital Era." Pp. 181-99 in *Critical Perspectives on Social Media and Protest: Between Control and Emancipation*, edited by L. Dencik and O. Leistert. London, England: Rowman & Littlefield.

Putnam, Robert D. 2001. *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon & Schuster.

Ram, Uri and Dani Filc. 2016. "The 14th of July of Daphni Leef: Class and Social Protest in Israel." *Capital & Class* 41:61-90.

Razin, Eran and Igal Charney. 2015. "Metropolitan Dynamics in Israel: An Emerging 'Metropolitan Island State'?" *Urban Geography* 36:1131-48.

Rosen, Ori, Wenxin Jiang, Gary King, and Martin A. Tanner. 2001. "Bayesian and Frequentist Inference for Ecological Inference: The R×C Case." *Statistica Neerlandica* 55:134-56.

Rosenhek, Zeev and Michael Shalev. 2014. "The Political Economy of Israel's 'Social Justice' Protests: A Class and Generational Analysis." *Contemporary Social Science* 9:31-48.

Royston, Patrick and Willi Sauerbrei. 2008. *Multivariable Model-building: A Pragmatic Approach to Regression Analysis based on Fractional Polynomials for Modelling Continuous Variables*. Hoboken, NJ: John Wiley.

Rüdig, Wolfgang and Georgios Karyotis. 2014. "Who Protests in Greece? Mass Opposition to Austerity." *British Journal of Political Science* 44:487-513.

Sanches, Pedro, Eric-Oluf Svee, Markus Bylund, Benjamin Hirsch, and Magnus Boman. 2014. "Knowing Your Population: Privacy-sensitive Mining of Massive Data." arXiv preprint arXiv:1412.2247.

Sasley, Brent E. and Harold M. Waller. 2016. *Politics in Israel: Governing a Complex Society*. Oxford: Oxford University Press.

Sauerbrei, Willi and Patrick Royston. 2016. "Multivariable Fractional Polynomial Models." in *Wiley Statsref: Statistics Reference Online*, edited by N. Balakrishnan, T. Colton, B. Everitt, W. Piegorsch, F. Ruggeri, and Teugels. doi:10.1002/9781118445112.stat07861.

Shalev, Michael. 2007. “The Welfare State Consensus in Israel: Placing Class Politics in Context.” Pp. 193-213 in *The Welfare State, Legitimacy and Social Justice*, edited by S. Mau and B. Veghte. Aldershot, England: Ashgate.

Shamir, Michal. 2015. “Introduction: The 2013 Elections and Israeli Democracy.” Pp. 1-16 in *The Elections in Israel, 2013*, edited by M. Shamir. New Brunswick, NJ: Transaction.

Shamir, Michal and Asher Arian. 1999. “Collective Identity and Electoral Competition in Israel.” *American Political Science Review* 93:265-77.

Shamir, Michal, Shira Dvir-Gvirzman, and Rafael Ventura. 2017. “Taken Captive by the Collective Identity Cleavage: Right and Left in the 2015 Elections.” Pp. 139-64 in *The Elections in Israel 2015*, edited by M. Shamir and G. Rahat. New Brunswick, NJ: Transaction.

Sloan, Luke and Jeffrey Morgan. 2015. “Who Tweets with Their Location? Understanding the Relationship between Demographic Characteristics and the Use of Geoservices and Geotagging on Twitter.” *PLoS One* 10: e0142209.

Smoreda, Zbigniew, Ana-Maria Olteanu-Raimond, and Thomas Couronné. 2013. “Spatiotemporal Data from Mobile Phones for Personal Mobility Assessment.” Pp. 745-69 in *Transport Survey Methods: Best Practice for Decision Making*, edited by J. Zmud, M. Lee-Gosselin, M. Munizaga, and J. A. Carrasco. Bingley, England: Emerald Group.

Sobolev, Anton, Keith Chen, Jungseock Joo, and Zachary C. Steinert-Threlkeld. 2019. “News and Geolocated Social Media Accurately Measure Protest Size.” Unpublished Working Paper, June 2019.

Soule, Sarah A. 2004. “Diffusion Processes within and across Movements.” Pp. 294-310 in *The Blackwell Companion to Social Movements*, edited by D. A. Snow, S. A. Soule, and H. Kriesi. Oxford, England: Blackwell.

Staggenborg, Suzanne and Josée Lecomte. 2009. “Social Movement Campaigns: Mobilization and Outcomes in the Montreal Women’s Movement Community.” *Mobilization* 14:163-80.

Steinert-Threlkeld, Zachary C. 2017. “Spontaneous Collective Action: Peripheral Mobilization during the Arab Spring.” *American Political Science Review* 111: 379-403.

Steinert-Threlkeld, Zachary C. 2018. *Twitter as Data*. Cambridge, MA: Cambridge University Press.

Talshir, Gayil. 2015. “‘The New Israelis’: From Social Protest to Political Parties.” Pp. 31-58 in *The Elections in Israel, 2013*, edited by M. Shamir. New Brunswick, NJ: Transaction.

Tarrow, Sidney G. 2011. “Cycles of Contention.” Pp. 195-214 in *Power in Movement: Social Movements and Contentious Politics* (3rd ed.). Cambridge, England: Cambridge University Press.

Tilly, Charles and Sidney G. Tarrow. 2007. *Contentious Politics*. Boulder, CO: Paradigm.

Traag, V. A., R. Quax, and P. M. A. Sloot. 2017. “Modelling the Distance Impedance of Protest Attendance.” *Physica A: Statistical Mechanics and Its Applications* 468:171-82.

Tufekci, Zeynep. 2014. “Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls.” in *Proceedings of the Eighth International Conference on Weblogs and Social Media*. AAAI Press. Retrieved August 15, 2017 (http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/viewFile/8062/8151).

Valentino-Devries, Jennifer, Natasha Singer, Michael H. Keller, and Aaron Krolik 2018. “Your Apps Know Where You Were Last Night, and They’re Not Keeping It Secret.” *New York Times*. Retrieved December 10, 2018 (https://www.nytimes.com/interactive/2018/12/10/business/location-data-privacy-apps.html).

Van Aelst, Peter and Stefaan Walgrave. 2001. “Who Is That (Wo)man in the Street? From the Normalisation of Protest to the Normalisation of the Protester.” *European Journal of Political Research* 39:461-86.

Verba, Sidney, Kay Lehman Schlozman, and Henry E. Brady. 1995. *Voice and Equality: Civic Voluntarism in American Politics*. Cambridge, England: Harvard University Press.

von Mörner, Moritz. 2017. “Application of Call Detail Records—Chances and Obstacles.” *Transportation Research Procedia* 25:2233-41.

Walgrave, Stefaan and Bert Klandermans. 2010. “Open and Closed Mobilization Patterns: The Role of Channels and Ties.” Pp. 169-93 in *The World Says No to War: Demonstrations against the War on Iraq*, edited by S. Walgrave and D. Rucht. Minneapolis: University of Minnesota Press.

Walgrave, Stefaan and Joris Verhulst. 2006. “Towards ‘New Emotional Movements’? A Comparative Exploration into a Specific Movement Type.” *Social Movement Studies* 5:275-304.

Walgrave, Stefaan and Joris Verhulst. 2008. “Protest Surveying: Testing the Feasibility and Reliability of an Innovative Methodological Approach to Political Protest.” PARTIREP Working Paper. Retrieved May 30, 2016 (http://www.partirep.eu/sites/default/files/working_paper/document/1236065530.pdf).

Walgrave, Stefaan and Joris Verhulst. 2009. “Government Stance and Internal Diversity of Protest: A Comparative Study of Protest against the War in Iraq in Eight Countries.” *Social Forces* 87(3):1355-1387.

Walgrave, Stefaan and Joris Verhulst. 2011. “Selection and Response Bias in Protest Surveys.” *Mobilization* 16:203-22.

Walgrave, Stefaan, Ruud Wouters, and Pauline Ketelaars. 2016. "Response Problems in the Protest Survey Design: Evidence from Fifty-one Protest Events in Seven Countries." *Mobilization* 21:83-104.

Wallace, Tim and Alicia Parlapiano. . 2017. "Crowd Scientists Say Women's March in Washington Had 3 Times More People than Trump's Inauguration." *The New York Times*. Retrieved January 23, 2017 (https://www.nytimes.com/interactive/2017/01/22/us/politics/womensmarchtrumpcrowdestimates.html).

Yoruk, Erdem and Murat Yuksel. 2014. "Class and Politics in Turkey's Gezi Protests." *New Left Review* 89:103-23.

Zickuhr, Kathryn. 2013. "Location-based Services." *Pew Research Center*. Retrieved October 15, 2018 (http://pewinternet.org/Reports/2013/Location.aspx).

Zurbarán, Mayra, Pedro M. Wightman, Liliana Gonzalez, and Miguel Labrador. 2014. "A Survey on Privacy in Location-based Services." *Ingeniería y Desarrollo* 32:314-43.

Author Biographies

Assaf Rotman is a postdoctoral researcher at the Department of Sociology and Anthropology, Tel Aviv University. His research interests include different forms of income and educational inequality, social movements and distributive justice.

Michael Shalev is professor emeritus of Sociology and Political Science at the Hebrew University of Jerusalem. His research, much of it on Israel, focuses on political economy, stratification and the politics of inequality.